ENGINEERING OF TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES (TALENS) FOR TARGETED GENOME EDITING

BY

NING SUN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Huimin Zhao, Chair
Professor Susan A. Martinis
Associate Professor Fei Wang
Professor Yi Lu
Professor James H. Morrissey

ABSTRACT

In the post-genome era, one of the most important topics of research is to edit or program genomic sequences and to generate desired phenotypes. Although virus-based strategies have long been developed to for efficient gene insertion, the random or semi-random integration can disrupt certain endogenous genes and cause unpredictable phenotypes. In contrast, targeted genome editing enables researchers to tailor genomic loci in a specific manner. Applications include studying gene functions, engineering microbes for industrial fermentation, improving traits in crop plants and livestock, treating human diseases, etc. This thesis describes my efforts on engineering transcription activator-like effector (TALE) nucleases (TALENs) as an efficient tool for targeted genome editing.

Targeted genome engineering relies on the introduction of a site-specific double-strand break (DSB) in a pre-determined genomic locus by a rare-cutting DNA endonuclease. Subsequent repair of this DSB by non-homologous end joining or homologous recombination generates the desired genetic modifications such as gene disruption, gene insertion, gene correction, etc. For this purpose, I have constructed TALEN architecture by fusing the DNA binding domain of TALE and a FokI non-specific DNA cleavage domain. TALEs are isolated from the plant pathogenic bacteria from the genus Xanthomonas and their DNA binding domains are composed of a series of tandem repeats. Each repeat comprises 33–35 amino acids and recognizes a single nucleotide. The DNA recognition specificity is conferred by the highly variable amino acids at positions 12 and 13 (e.g., NI recognizes adenine, HD recognizes cytosine, NG recognizes thymine, and NN recognizes guanine and adenine). This simple code and

independent DNA binding of the repeat units enable TALEs to bind to any custom-designed DNA sequence. The fusion of a FokI cleavage domain makes TALENs a new class of artificial DNA endonucleases which serve as a powerful tool for targeted genome editing.

To monitor *in vivo* activities of TALEN, I have constructed various reporter systems constructed in yeast and human cells. To maximize the genome editing efficiency of TALENs, I have optimized the TALEN scaffold by the truncation of N- and C-termini of TALEs. Two TALEN scaffolds were identified with efficient activity in modifying both yeast and human genomes. To further improve TALEN platform, I have constructed a high-throughput screening system and identified the SunnyTALEN architecture through directed evolution. Compared with the existing TALEN platform, SunnyTALEN shows significantly increased genome editing efficacy in both yeast and human cells. To demonstrate the application of TALEN technology in human therapeutics, I have corrected the sickle cell disease mutation in patient-derived induced pluripotent stem cells. The corrected stem cells can serve as a regenerative medicine for the treatment of human genetic disorders. Lastly, I have created a novel single-chain TALEN architecture, which can be used to decrease the payload for efficient TALEN delivery.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Professor Huimin Zhao, for his continuous support and encouragement to my research projects and career development. His intelligence, diligence and love in research demonstrate what a real "USTCer" should be like and he has been and will always be my role model.

I would like to thank Professor Fei Wang for kindly sharing the expertise for handling human stem cells. I am very grateful for the encouragement and the great suggestions that I received from my thesis committee: Professor Susan Martinis, Professor Yi Lu, Professor Fei Wang and Professor James Morrissey. I would also like to thank Dr. Ben Montez and Dr. Barbara Pilas at the Roy J. Carver Biotechnology Center for their assistance with cell sorting.

I would like to thank all of the members of the Zhao lab. Their good cheer and humor made the Zhao group a big happy family. I am grateful to Dr. Fei Wen for mentoring me when I did rotation in the lab. Thank her and Dr. Zengyi Shao for their valuable suggestions on my career development. I would like to thank Jing Liang, Zehua Bao and Xiong Xiong for their assistance with my research projects. Special thanks to Emmanuel "Luigi" Chanco, Tong "Tony" Si and Sujit Jagtap for their friendship. The lunch parties we had were full of joy and a lot of excellent ideas came from the random discussions.

Last but not least, I would like to thank my parents for their love and support over the years. Special thanks to my wife, Dr. Mianzhi Gu, for being with me through the ups and downs of graduate school. I could not have enjoyed this journey without her.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

1.1. Targeted genome editing

Targeted genome engineering or editing enables researchers to modify genomic loci of interest in a precise manner, which has various applications in industry, agriculture and human therapeutics. Virus-based strategies have been developed to efficiently integrate an exogenous gene into a mammalian genome (1). However, the integration process is often random, which may disrupt certain endogenous genes and cause unpredictable phenotypes. For example, in the case of human gene therapy, insertional mutagenesis caused by random integration of viral vectors is a major safety concern, which limits the wide adoption of this technology (2,3). Chimeric zinc finger recombinases (4) and zinc finger transposases (5) have been developed to integrate foreign DNA into the genome in a more specific manner, but they are still in the preliminary stage and have met limited success. A more established strategy for targeted genome engineering relies on the introduction of a site-specific double-strand break (DSB) in a pre-determined locus by an engineered DNA endonuclease. Subsequent repair of this DSB by non-homologous end joining (NHEJ) or homologous recombination (HR) will generate desired genetic modifications (6). The NHEJ mechanism can disrupt a gene by introducing frame-shift mutations while the HR mechanism will result in gene deletion, gene insertion or gene correction (**Figure 1.1**).

For an engineered DNA endonuclease to be widely used in targeted genome engineering, two criteria must be met: i) it must recognize a long DNA sequence with high specificity in order to avoid cytotoxic off-target DNA cleavage; ii) it must be readily

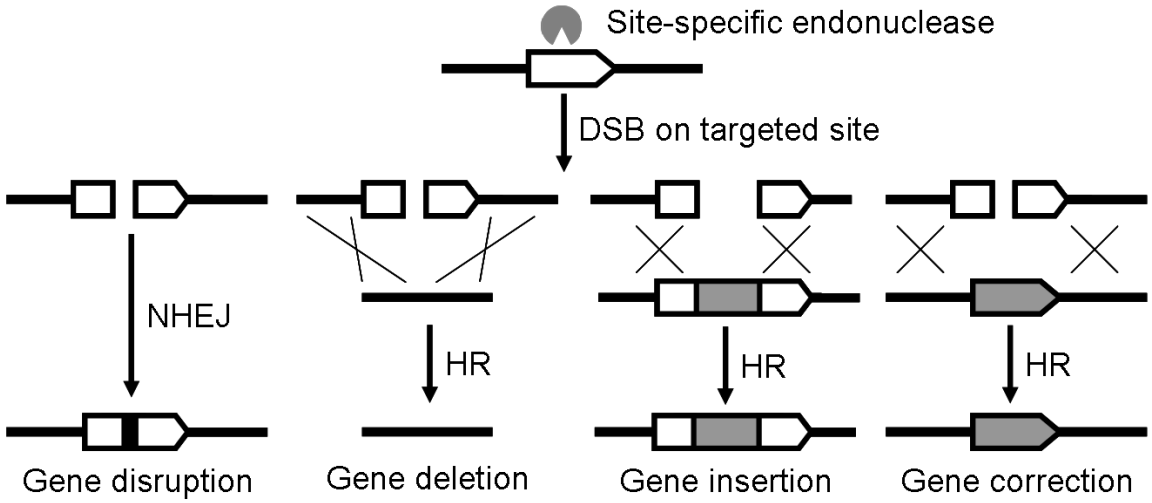

Figure 1.1. The DNA DSB generated by a site-specific endonuclease will be repaired by either NHEJ or HR, resulting in gene disruption, gene deletion, gene addition or gene correction.

designed to recognize and cleave a defined sequence in the genome. Three major classes of DNA endonucleases have been developed as genome engineering tools, including engineered homing endonucleases (also called meganucleases), zinc finger nucleases, and transcription activator-like effector nucleases (TALENs).

1.2. Engineered homing endonucleases

Homing endonucleases (HEs), also known as meganucleases, represent a family of naturally occurring rare-cutting endonucleases, which can be found in all kingdoms of life. HEs specifically recognize and cleave long DNA sequences (14-40 bp), which represent a promising tool to promote chromosomal DSBs for targeted genome engineering (reviewed by (7,8)). Among all the identified HE families, members of the eukaryal and archael LAGLIDADG family exhibit the highest overall DNA recognition specificity. LAGLIDADG homing endonucleases (LHEs) use antiparallel β-sheets as a

DNA recognition module by forming a saddle on the DNA helix major groove, and the DNA cleavage is catalyzed by the divalent metal cations in the active site (**Figure 1.2A**) (9). Some HE proteins such as I-CreI and I-MsoI form homodimers and cleave palindromic or pseudo-palindromic target sites, while other members such as I-SceI and I-AniI are monomeric, cleaving non-palindromic DNA sequences.

Several hundreds of HEs have been identified. However, the repertoire of DNA recognition sequences is very limited compared to the size of the human genome (>20,000 genes), making it nearly impossible to find a natural HE recognition site in a pre-determined genomic region. Therefore, HEs must be engineered to recognize and cleave target DNA sequences, which is a very challenging task due to the complex, highly cooperative network of protein-DNA contacts and the tight coupling of cognate site binding to subsequent catalysis of DNA bond cleavage (**Table 1.1**). Directed evolution strategies have been developed to modify the sequence specificity of known HEs (10,11). For example, Doyon and coworkers (12) and Chen and coworkers (13) successfully isolated I-SceI variants capable of cleaving novel DNA sequences by the help of a bacterial based *in vivo* selection system. Instead of creating or screening large protein libraries, the Baker group used computational protein design to reprogram HEs and successfully altered the specificity of I-MsoI and I-AniI (14-16). To engineer homodimeric HEs such as I-CreI to recognize novel non-palindromic targets, each monomer has to be engineered separately to target its palindromic site as the first step. Next, the two different I-CreI monomers can be co-expressed and heterodimerized upon binding the chimeric DNA sequence, resulting in cleavage (17). Using this strategy, however, cleavage-competent I-CreI homodimers may also form and cleave their

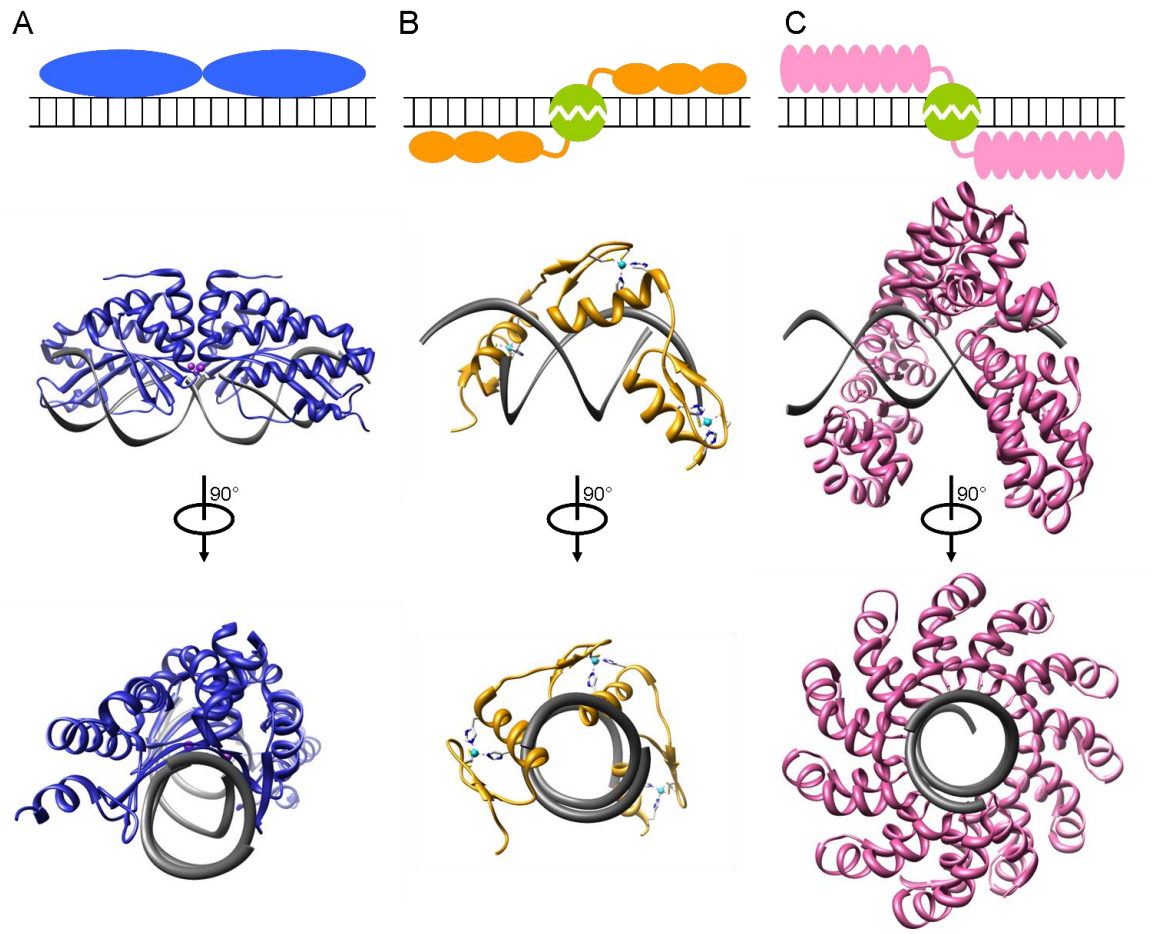

Figure 1.2. Schematic and structures of engineered site-specific endonucleases. Two perpendicular views are presented, with the DNA duplex shown in grey. (**A**) Homing endonuclease I-CreI (blue) binds to the target site as a dimer. Antiparallel β-sheets are used for DNA recognition and DNA cleavage is catalyzed by the magnesium ions (purple spheres) within the central catalytic core (adapted from (18), PDB no. 2VBL). (**B**) Zinc finger nucleases function as heterodimers and bind to DNA targets by zinc finger domains (orange). DNA cleavage is executed by the dimerization of FokI DNA cleavage domain (green). Zinc fingers coordinate zinc ions (cyan spheres) with a combination of cysteine and histidine residues. The α-helix of each zinc-finger interacts with three nucleotides in the DNA major groove (adapted from (19), PDB no. 1MEY). (**C**) TAL effector nucleases use central tandem repeats (pink) for DNA sequence recognition and the FokI DNA cleavage domain (green) for DNA cleavage. The TAL tandem repeats form a right-handed super-helical structure, binding to the major groove of DNA (adapted from (20), PDB no. 3V6T).

Table 1.1. Comparison of three tailored site-specific endonucleases.

	Homing endonuclease	**Zinc finger nuclease**	**TAL effector nuclease**
Typical length of recognition site	14-40bp	18-24 bp with 5-7bp spacer	30-50 bp with 10-30 bp spacer
Typical protein size	<40 kD	<30 kD*	>100 kD[a]
Modularity	low	medium	high
Advantages	1. non-specific FokI DNA cleavage domain is not required 2. High DNA cleavage efficiency	1. relatively easy to tailor the substrate specifcity 2. well characterized in terms of specificity, affinity and toxicity	1. easy to tailor the substrate specifcity 2. the length of the recognition site can be freely adjusted
Disadvantages	1. difficult to tailor the substrate specificity 2. specificity and toxicity have not been determined systematically	1. sequence bias; prefer G rich sequence 2. some ZFNs exhibit off-target cleavage and cytotoxicity	1. no code for recognizing guanine specifically 2. specificity and toxicity have not been determined systematically 3. large protein size may cause difficulty in delivery

[a] the protein size of each monomer

respective palindromic sequences, which can limit the safety or efficacy via off-target cleavages. With the help of computational protein design, Fajardo-Sanchez and coworkers (21) re-designed the interaction surface of I-CreI to obtain an obligate heterodimer, which prefers a single chimeric DNA sequence and prevents the formation of I-CreI homodimers. For a similar reason, single-chain I-CreI was created by inserting a polypeptide linker between two copies of the I-CreI genes (22-24). Based on the single-chain architecture, Ulge and coworkers (25) used computational design to generate thousands of monomeric I-CreI variants with altered specificities at 16 different base pair positions in the 22 bp target site. The single-chain design not only decreased the rate of off-target cleavage and cytotoxicity, but also simplified the engineering process.

Among all HEs, I-CreI is the first one whose DNA recognition specificity was successfully modified to target a defined sequence within the human genome. Alteration of local substrate specificity of DNA-binding subdomains may not perturb the overall

protein structure and function, making I-CreI certain degree of modularity. By combining rational design and high throughput screening, I-CreI variants were engineered to target the human *RAG1* gene associated with severe combined immunodeficiency (26,27) and the human *XPC* gene associated with xeroderma pigmentosum (18,28). Similarly, several I-CreI variants were generated to cleave Herpes simplex virus type 1 (HSV1) genome sequences (29). Moreover, single-chain I-CreI was used to induce high levels of gene targeting at the endogenous human *RAG1* gene locus (27,30). In addition, Takeuchi and coworkers (31) used I-OnuI as a template and created a variant to cleave a sequence within the monoamine oxidase B (*MAO-B*) gene in human cells. Since this gene is a potential therapeutic target for a variety of neurodegenerative disorders, targeted disruption or modifications of the *MAO-B* gene might have great value for future clinical research.

1.3. Zinc finger nucleases

Zinc finger nucleases (ZFNs) are artificial DNA nucleases constructed by fusing several zinc finger domains to the sequence-independent cleavage domain of the type IIS restriction endonuclease FokI (32). Zinc finger domains serve as DNA recognition modules, and modifying the zinc finger domains can target the ZFNs to novel DNA sequences. A typical zinc finger domain contains an α-helix recognizing a specific DNA triplet. Therefore, a typical ZFN monomer containing 3 or 4 zinc finger domains recognizes a 9 or 12 bp DNA target site. Since the non-specific DNA cleavage domain of FokI functions as a dimer and has the highest cleavage efficiency in the inverse orientation (33), the typical custom-designed ZFN functions as a heterodimer recognizing

an 18-24 bp DNA sequence with a 4-6 bp spacer between each half site (**Figure 1.2B**). Compared with the engineered homing endonucleases, custom-designed ZFNs are easier to construct due to their moderate degree of modularity (**Table 1.1**). Multiple platforms such as the CompoZr^{TM} service, modular assembly, Oligomerized Pool Engineering (OPEN), and Context Dependent Assembly (CoDA) have been used to generate site-specific ZFNs (reviewed by (34-36)). Because several zinc finger domains can tolerate mismatched DNA bases, off-target ZFN cleavage has been reported (37,38). The recognition of secondary degenerate sites of ZFNs will generate undesired genomic DSBs and result in unexpected side effects. Therefore, the sequence fidelity of ZFNs has to be improved for future ZFN design because it is critical to their safe and successful application in targeted genome engineering (**Table 1.1**). Recent advances in the ZFN technology involve a large number of applications such as gene disruption, gene insertion, gene correction, and chromosomal rearrangement (**Figure 1.1**), as will be highlighted below.

1.3.1 Gene disruption

The errors introduced during NHEJ-mediated chromosomal DSB repair can be used to achieve gene disruption (**Figure 1.1**). This approach has been applied in various mammalian cells to efficiently knock out targeted genes in a single step (recently reviewed by (36)). There are ongoing clinical trials for the treatment of human immunodeficiency virus (HIV) infection, which is based on ZFN-mediated *CCR5* gene disruption (39,40). Recently, Cradick and coworkers (41) described a novel therapeutic strategy for treatment of hepatitis B by using ZFNs to target the hepatitis B virus genome.

Besides human cells, ZFNs were successfully applied to other mammals such as rat (42,43), rabbit (44) and pig (45). Moreover, in addition to NHEJ-mediated gene disruptions, Zou and coworkers inactivated the endogenous *PIG-A* locus associated with paroxysmal nocturnal hemoglobinuria in human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) by DSB-induced HR (46). Successful HR-mediated targeting resulted in the insertion of an antibiotic selection cassette into the *PIG-A* gene exon, thereby inactivating the *PIG-A* gene.

1.3.2 Gene insertion

Precise gene insertion can be achieved by ZFN-induced DNA repair (**Figure 1.1**). Hockemeyer and coworkers (47) reported the highly efficient targeting of three genes in human iPSCs via ZFN-based HR-mediated genome insertion. Specifically, they successfully integrated the enhanced green fluorescent protein (eGFP) reporter gene into the *OCT4*, *AAVS1* and *PITX3* loci. Note that the *AAVS1* locus within the *PPP1R12C* gene on chromosome 19 is considered as a 'safe harbor' for addition of a transgene into the human genome (48). DeKelver and coworkers (49) described the use of ZFNs to insert various expression cassettes into the *AAVS1* locus for isogenic transgenesis application. ZFN-mediated gene insertion turned out to be efficient (at a frequency of up to 15%) in both transformed cell lines (K562, HeLa, HEK293, U2OS, and others) and primary human cells (fibroblasts and ESCs). It was demonstrated that ZFN-driven gene addition into the *AAVS1* locus did not introduce any adverse effect on the cell resulting from its disruption. Moreover, both promoterless (driven by the native *PPP1R12C* gene promoter) and promoter-containing inserts placed into the *AAVS1* locus exhibited consistent gene

expression levels over 50 cell generations. Later, Zou and coworkers (50) integrated a single-copy *gp91*(*phox*) gene into the *AAVS1* locus for treatment of the X-linked chronic granulomatous disease (X-CGD). X-CGD patients have a defect in neutrophil microbicidal reactive oxygen species (ROS) generation because of the *gp91*(*phox*) deficiency. With the help of a ZFN-induced DSB, a therapeutic *gp91*(*phox*) minigene was inserted into one allele of the *AAVS1* locus in the X-CGD iPSCs without off-target insertions, which complemented the *gp91*(*phox*) gene deficiency and substantially restored neutrophil ROS production.

1.3.3 Gene correction

Gene correction is mediated by DSB-induced HR, which can be exploited for gene replacement, especially for the treatment of monogenic diseases. Compared with gene insertion, *in situ* gene correction is more challenging because the designed ZFN cleavage site must be close to or directly at the site of the mutation (**Figure 1.1**). Through site-specific chromosomal DSBs induced by ZFNs, Sebastiano and coworkers (51) demonstrated HR-mediated efficient correction of the sickle cell anemia mutation in patient-derived human iPSCs. Two pairs of ZFNs were constructed by the publicly available OPEN method (52) to target the disease-causing β-globin gene. They first used those ZFNs to simultaneously correct the E6V mutation and insert a drug-resistance cassette into a neighboring intron 58 or 82 bp downstream of the ZFN cleavage sites. Isolation of drug-resistant clones and PCR screening followed by DNA sequencing enabled them to identify several iPSC clones containing the successfully corrected allele. Next they used the Cre recombinase to excise the selection gene cassette flanked by the

loxP sites, leaving a residual 34 bp *loxP* site "scar" in the intron of the corrected *β*-globin gene. Around the same time, Zou and coworkers (53) achieved site-specific gene correction of the *β*-globin gene in patient-derived human iPSCs. The ZFNs they used were custom designed by CompoZr^{TM} from Sigma-Aldrich (St. Louis, MO, USA). A similar two-step strategy was used to correct the sickle cell disease causing mutation: (i) Correction of the E6V mutation by ZFN-induced HR and integration of a *loxP*-flanked drug-resistant cassette into a neighboring intron; (ii) Cre recombinase-mediated excision of the drug-resistant cassette. To evaluate the gene expression of the corrected *β-globin* gene, they differentiated the corrected iPSCs into erythroid cells and surprisingly discovered that the expression level of the corrected *β*-globin gene was partially repressed. Since Cre-recombinase mediated gene excision in step (ii) left a residual *loxP* sequence in the intron, it is possible that the remaining *loxP* "scar" interfered with the transcriptional regulatory elements and thus altered the gene expression from the targeted allele. In addition to the unpredicted effect from the residual *loxP* site, such a strategy may not be applicable to all genes, especially when the location of the desired editing events is away from the intron region.

Instead of using the Cre/*loxP* recombination system that leaves small ectopic sequences in the targeted genome, Yusa and coworkers combined ZFN with the *piggyBac* technology and achieved a "scarless" gene correction in patient-derived iPSCs (54). *piggyBac* is a moth-derived DNA transposon, which enables the removal of transgenes flanked by its inverted repeats without leaving any residual sequences (55). To correct the single mutation (E342K) in the *A1AT* gene causing α_1-antitrypsin deficiency, ZFN pairs were designed to specifically cleave the mutation site. Through ZFN-induced HR, they

corrected the *A1AT* gene mutation and simultaneously inserted a *PGK-puroΔtk* cassette flanked by *piggyBac* repeats into the TTAA site. Targeted clones were isolated by puromycin selection. Next, the selection cassette was excised from the iPSC genome by transient expression of the *piggyBac* transposase and subsequent counter-selection, leaving no residual ectopic sequences in the targeted allele. Not only did their approach achieve seamless gene correction of diseased human iPSCs, but also enabled biallelic correction with high efficiency. Without using a selectable marker, Soldner and coworkers employed a selection-free targeting strategy (56). Briefly, given the high gene editing activity of the ZFNs, they constructed a donor vector lacking a selection cassette, consisting of only a homology arm flanking the ZFN cleavage site carrying a wild-type α-synuclein gene. An eGFP-expressing plasmid was coelectroporated into Parkinson's disease patient-derived iPSCs together with the donor construct and ZFN-expression vectors to enrich transfected cells by fluorescence-activated cell sorting. One correctly targeted clone with the A53T (G209) mutation repaired was isolated after screening 240 single-cell-derived clones by Southern blot analysis. The high efficiency suggests that one could isolate corrected clones by dilution cloning without going through drug selection.

1.3.4 Chromosomal rearrangements

Chromosomal rearrangements include large-scale gene deletions, insertions, duplications and inversions, which are associated with many genetic diseases and cancer (57). ZFNs have been utilized to generate targeted chromosomal rearrangements, which enables researchers to study gene functions at the genomic level (**Figure 1.3**). Brunet and

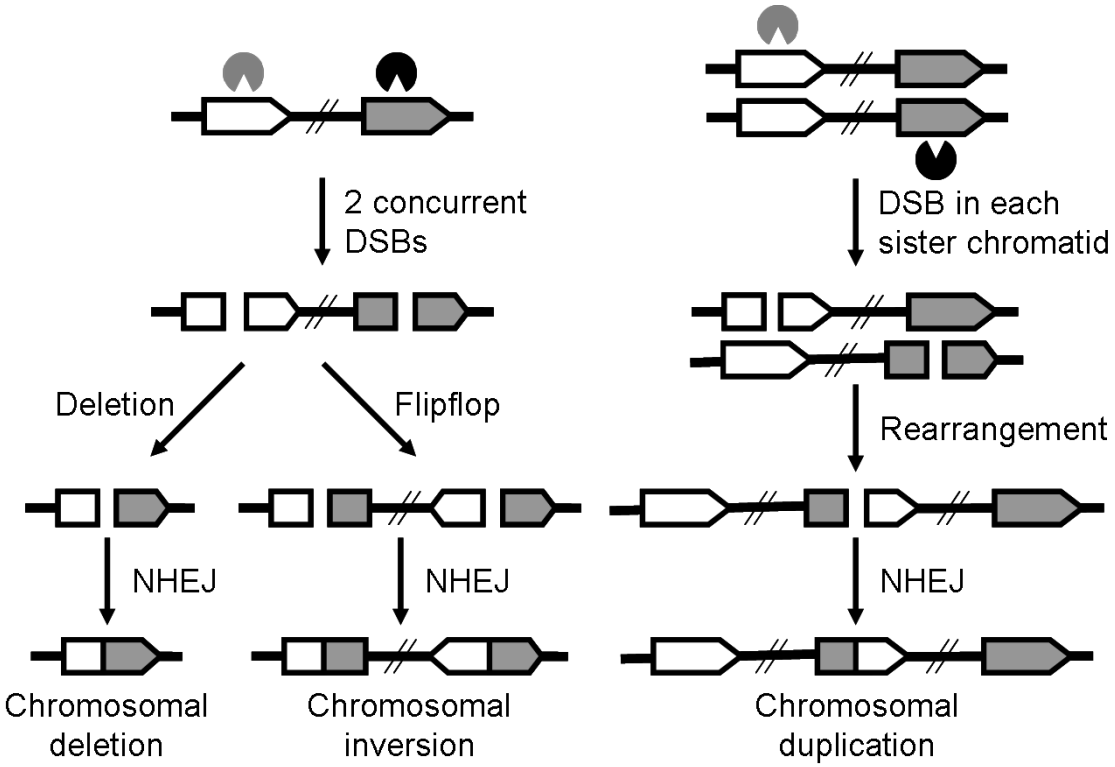

Figure 1.3. The concurrent DSBs generated by two endonucleases on different chromosomal loci will be ligated by the NHEJ mechanism, resulting in chromosomal deletion, inversion or duplication.

coworkers (58) simultaneously introduced two pairs of ZFNs recognizing the *PPP1R12* gene on chromosome 19 and the *IL2Rγ* gene on the X chromosome into human cells, respectively. The concurrent DSBs at the two endogenous loci on different chromosomes induced chromosomal translocations between chromosome 19 and the X chromosome via NHEJ-mediated DSB repair. Using a similar approach, Lee and coworkers (59) used engineered ZFNs to generate targeted deletions of genomic segments in human cells. They designed two ZFN pairs to generate two concurrent DSBs in the same chromosome (**Figure 1.3**). The breakpoint joined together by endogenous NHEJ resulted in targeted deletions of the genomic segment between the two ZFN cleavage sites. By applying this method to the HEK293T cell line, they successfully deleted predetermined genomic DNA

segments in the range of ~729 bp to 15 Mb with frequencies of 0.1% to 10%. In addition to genomic deletions, they demonstrated that two concurrent DSBs introduced by ZFN pairs were sufficient to promote frequent genomic inversions and duplications in human cells with frequencies ranging from 0.01% to 5% (60). Harnessing various combinations of two ZFNs, they achieved duplications of genomic DNA between the two cleavage sites whose length ranged from 230 kb to 835 kb (**Figure 1.3**). They also demonstrated that the two concurrent genomic DSBs from ZFN cleavages induced inversions of 15 kb to 15 Mb DNA segments in the human genome (**Figure 1.3**). As proof of concept for therapeutic application of this technique, they constructed a ZFN pair to target the intron 1 homolog in the human *F8* gene, whose inversion causes severe hemophilia A (61). The genomic DSB generated by the ZFN pair induced the inversion of the 140 kb DNA segment bearing the promoter and exon 1 of the *F8* gene with a frequency of 0.2%-0.4%. Their strategy demonstrated the promise of restoring genomic integrity in severe hemophilia A patients by reverting the inverted DNA segment back to the wild-type orientation.

1.4. TALENs

Even though ZFNs have been used for targeted genome editing in various organisms, two major limitations prevent their wider applications. ZF domains have limited modularity due to the context-dependent DNA-binding effects, making it difficult for ZFNs to target any desired DNA sequence (62). Moreover, lack of specificity of some ZF domains can generate off-target cleavage, leading to undesired mutations and chromosomal aberrations (38,63). Recently, TALENs have rapidly emerged as an alternative genome

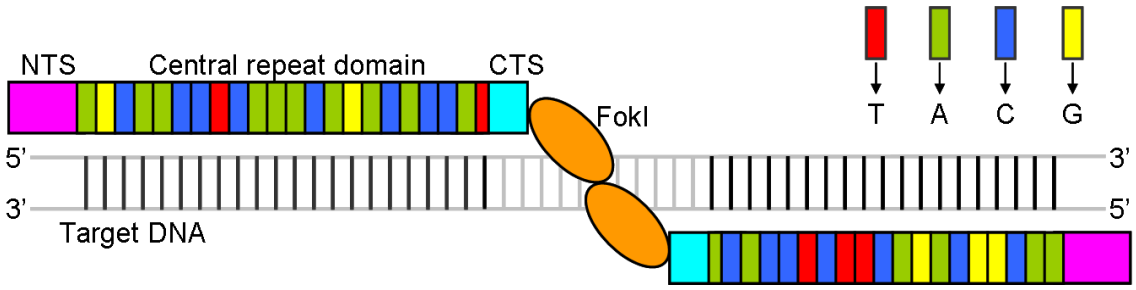

Figure 1.4. Schematic of TALEN architecture. A TALEN is composed of a NTS (pink box), a central repeat domain, a CTS (cyan box) and a FokI catalytic domain (orange oval). The central repeat domain comprises a series of repeat units that are responsible for specific recognition of thymine (red boxes), adenine (green boxes), cytosine (blue boxes) and guanine (yellow boxes). The formation of a heterodimer by two TALENs in a tail-to-tail orientation at the target site executes a site-specific DNA DSB. The TALE binding sites on the target DNA are shown in black and the spacer is shown in grey.

editing tool to ZFNs (recently reviewed by (64)). Similar to ZFNs, TALENs use the non-specific FokI domain as the DNA cleavage module and function as dimmers (**Figure 1.2C** and **Figure 1.4**). However, the DNA binding domains of TALENs are composed of a series of tandem repeats as in TALEs of the plant pathogenic bacteria from the genus *Xanthomonas* (as reviewed in (65,66)). Each repeat comprises 33-35 aa and recognizes a single nucleotide. The last repeat typically has only 20 aa, and is therefore called a 'half-repeat'. The DNA recognition specificity is conferred by the highly variable amino acids at positions 12 and 13 (*e.g.* NI recognizes adenine, HD recognizes cytosine, NG recognizes thymine, and NN recognizes guanine and adenine) (67,68). Unlike the context-dependent DNA binding of ZFNs, TALENs can be easily and rapidly constructed to target essentially any DNA sequence due to the simple protein-DNA code and the modular nature. In addition, TALENs exhibit significantly reduced off-target effects and cytotoxicities compared with ZFNs, making them an efficient genome editing tool

(69,70). Within the last three years, TALENs have been widely applied to modify endogenous genes in a variety of organisms (**Table 1.2**). Applications include studying gene functions in model organisms, improving traits in crop plants and livestock, generating disease models, and treating genetic disorders in humans. In this article, we provide a comprehensive review of TALEN technology including the optimization of the scaffold, improvement of the DNA recognition specificity, and assembly of TALE repeat arrays. Due to the ease of design and high efficiency of genome editing, TALENs have opened up many new avenues for basic and applied biological research.

1.4.1. Scaffold optimization

The original TALEN construction was reported by two independent groups. Li and coworkers fused the full-length natural TALE (AvrXa7 and PthXo1) with the FokI catalytic domain, creating the TALENs bearing a 288 aa N-terminal segment (NTS) and a 295 aa C-terminal segment (CTS) (71). Based on a yeast reporter assay, the optimal spacer length between the two TALEN binding sites was determined to be 16-31 bp (**Table 1.3**). Alternatively, the TALENs created by Christian and coworkers encompasses a 287 aa NTS and a 231 aa CTS (72). This scaffold allows efficient DNA cleavage against target sites with 13-30 bp spacers (**Table 1.3**). Since naturally-occurring TALEs are transcription activators from a plant bacterial pathogen, their NTSs harbor protein secretion signal peptides while their CTSs contain nuclear localization signal peptides and a transcription activator domain (73). These sequences can impair the catalytic activity when fused with the FokI cleavage domain. To identify the optimum TALEN architecture with highest cleavage efficiency and minimal peptide portion, scaffold

Table 1.2. Applications of TALENs for targeted genome editing in various organisms.

Organisms	Genes	Ref.
Arabidopsis thaliana	*ADH1*	(74)
Brachypodium	*BdABA1, BdCKX2, BdCOI1, BdHTA1, BdRHT, BdSBP, BdSMC6, BdSPL*	(75)
Cattle (*Bos taurus*)	*ACAN, GDF8, GGTA, PRNP*	(76)
Cricket (*Gryllus imaculatus*)	*Gb'lac2*	(77)
Frog (*Xenopus tropicalis*)	*ets1, foxd3, grp78/bip, hhex, noggin, ptf1a/p48, sox9, tyr, vpp1*	(78,79)
Fruitfly (*Drosophila melanogaster*)	*CG9797, yellow*	(80)
Hamster (*Cricetulus griseus*)	*FUT8*	(81)
Human (*Homo sapiens*)	*ABL1, AKT2, ALK, ANGPTL3, APC, APOB, ATGL, ATM, AXIN2, BAX, BCL6, BMPR1A, BRCA1, BRCA2, C6orf106, CIITA, CBX3, CBX8, CCND1, CCR5, CDC73, CDK4, CELSR2, CFTR, CHD4, CHD7, CTNNB1, CYLD, DDB2, DDX60, DHX58, DHX9, DICER1, EIF2AK2, EIF2C2, ERCC2, EWSR1, EXT1, EXT2, EZH2, FANCA, FANCC, FANCF, FANCG, FES, FGFR1, FH, FLCN, FLT4, FOXO1, FOXO3, GLI1, GLUT4, HBB, HDAC1, HDAC2, HDAC6, HMGA2, HOXA13, HOXA9, HOXC13, HPRT1, IFI44L, IFIT1, IFIT2, IL2RG, JAK2, KRAS, LINC00116, MAOA, MAP2K4, MB21D1, MDM2, MET, MLH1, MSH2, MUTYH, MYC, MYCL1, MYCN, NAMPT, NBN, NCOR1, NCOR2, NDUFA9, NLRC5, NTF3, NUB1, OASL, OCT4, PDGFRA, PDGFRB, PHF11, PHF8, PITX3, PLA2G4A, PLIN1,PMS2, PPP1R12C, PTCH1, PTEN, QRICH1, RARA, RBBP5, RECQL4, RET, RIPK4, RTP4, RUNX1, SDHB, SDHC, SDHD, SETDB1, SIRT6, SMAD2, SORT1, SS18, STAT1, STAT6, SUZ12, TBK1, TFE3, TP53, TRIB1, TSC2, TTN, UNC93B1, VHL, XPA, XPC*	(69,70,74,82-90)
Medaka (*Oryzias latipes*)	*DJ-1*	(91)
Mouse (*Mus musculus*)	*Pibf1, Sepw1*	(92)
Nematode (*Caenorhabditis elegans*)	*ben-1*	(93)
Rat (*Rattus norvegicus*)	*BMPR2, IgM*	(94,95)
Rice (*Oryza sativa L.*)	*Os11N3, OsBADH2, OsCKX2, OsDEP1, OsSD1*	(75,96)
Silkworm (*Bombyx mori*)	*BmBlos2*	(97,98)
Swine (*Sus scrofa*)	*AMELY, DMD, GDF8, GGTA, GHRHR, IL2Rg, p65, RAG2, RELA, SRY*	(76)
Tobacco (*Nicotiana tabacum*)	*SurA, SurB*	(99)
Yeast *(Saccharomyces cerevisiae*)	*ADE2, LYS2, URA3*	(100)
Zebrafish (*Danio rerio*)	*aanat2, abcc9, adora1b, adora2aa, bmi1, cdh5, clc, crhr1, dip2a, elmo1, epas1b, fgf21, fh, golden, gpr103a, gria3a, hdc, hey2, hif1ab, ikzf1, jak3, moesina, myod, nmu, npy, phf6, pmch, pmchl, ponzr1, ppp1cab, prok2, prokr1, prokr2, ptpmt1, qrfp, ryr1a, ryr3, scl6a, tbx6, tgfa, tnikb, vip*	(101-107)

Table 1.3. Engineered TALEN scaffolds with different NTSs and CTSs.

NTS (aa)	CTS (aa)	Spacer (bp)	Reporter system	References
288	295	16-31	β-galactosidase assay in yeast	(71)
288	285	16[a]	transient expression assay in tobacco leaves	(108)
287	231	13-30	β-galactosidase assay in yeast	(72)
287	63	15[a]	mutagenesis in medaka embryos	(91)
207	63	14-32	β-galactosidase assay in yeast	(89)
207	31	10-16	β-galactosidase assay in yeast	(89)
153	47	12-21	dsEGFP assay in HEK293	(70)
153	17	12	dsEGFP assay in HEK293	(70)
136	63	12-20	Surveyor nuclease assay in K562	(84)
136	28	12-13	Surveyor nuclease assay in K562	(84)
136	18	13-16	β-galactosidase assay in yeast	(109)

[a] The spacer length was not optimized in the study

optimization has been carried out by several groups. Using a truncated NTS with only 136 aa, Miller and coworkers constructed a series of TALENs by trimming the CTS between the central repeat units and the FokI nuclease domain (84). This study found that different TALEN scaffolds prefer different spacer length between each TALEN binding site. One TALEN scaffold was identified bearing a 63 aa CTS, which could drive efficient gene modification in human cells when separated by 12-20 bp spacers (**Table 1.3**). This scaffold has since been widely applied for efficient genome editing in various species (79,82,85,93,94,102,105-107). In addition, another TALEN scaffold bearing only a 28 aa CTS with a narrower separation range of 12-13 bp was identified (**Table 1.3**). Later, Mussolino et al. and Christian et al. reported TALENs with even shorter CTSs that exhibit narrow optimal ranges of spacers (**Table 1.3**) (70,109). A systematic study was carried out by Sun and coworkers who constructed 10 different TALEN scaffolds with various NTSs and CTSs (89). The DNA cleavage activity of each scaffold was assayed against 10 substrates with different spacers in a yeast reporter system. Based on this

10×10 matrix, two TALEN scaffolds with high DNA cleavage efficiency in both yeast and human cells were identified. One bearing a 207 aa NTS and a 31 aa CTS prefers target sites with 10-16 bp spacers while another bearing a 207 aa NTS and a 63 aa CTS has highest efficiency when separated by 14-32 bp spacers (**Table 1.3**). It is noteworthy that the TALEN with a 50 aa NTS has no catalytic activity against any target sites. Later on, Gao and coworkers solved the crystal structure of the TALE NTS and discovered an extended N-terminal DNA binding region composed of the 127 aa immediately preceding the central repeat units (110). The 127 aa NTS features four continuous repeats. Each repeat contains two α-helices and an intervening loop (**Figure 1.5A**), a structural feature highly similar to that of the central repeat unit. Although the 127 aa NTS does not confer sequence specificity, it is crucial for DNA binding. This feature explains why all the effective TALEN scaffolds have at least 127 aa preceding the central repeat units (**Table 1.3**).

A second-generation GoldyTALEN scaffold has been demonstrated to improve the genome editing efficiency in zebrafish (101). GoldyTALEN is based on a scaffold previously reported by Miller and coworkers (84) which consists of a 136 aa NTS and a 63 aa CTS. However, it has 9 different aa substitutions at the NTS and 5 different aa substitutions at the CTS. Using the GoldyTALEN scaffold and zebrafish delivery system, certain loci were modified with 100% efficiency. Moreover, they provided the first example of HR-based genome editing in zebrafish using single-stranded DNA as a donor. The GoldyTALEN scaffold was also applied for efficient gene knockout in livestock (76).

Because the FokI catalytic domain must dimerize to become active, two TALEN subunits are assembled as heterodimers at the cleavage site. However, cleavage-

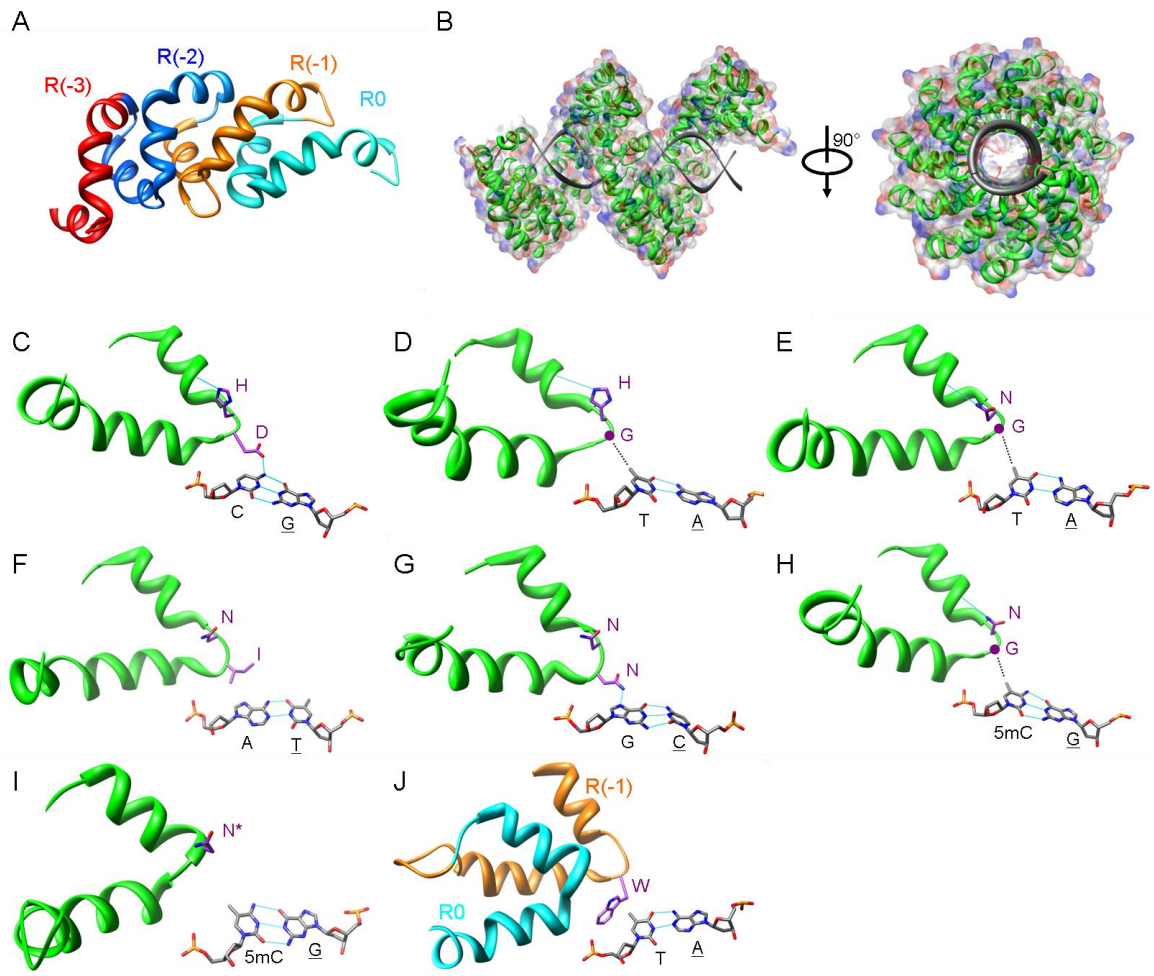

Figure 1.5. Crystal structures of TALEs. (**A**) The TALE NTS features four continuous repeats, which are important for the DNA binding affinity (adapted from (110)). (**B**) Overall structure of the TALE PthXo1 central repeat domain in complex with its target site (adapted from (111)). (**C**) Interaction of RVD HD with cytosine (adapted from (110)). Hydrogen bonds are indicated by cyan lines. (**D**) Interaction of RVD HG with thymine (adapted from (111)). A nonpolar van der Waals interaction is shown in a dotted line. (**E**) Interaction of RVD NG with thymine (adapted from (110)). (**F**) Interaction of RVD NI with adenine (adapted from (111)). (**G**) Interaction of RVD NN with guanine (adapted from (111)). (**H**) Interaction of RVD NG with 5-methyl cytosine (adapted from (112)). (**I**) Interaction of RVD N* with 5-methyl cytosine based on a structural model. (**J**) Interaction of the NTS with the 5'- preceding thymine. (Adapted from (111)). The coloring of the two repeats in TALE NTS matches that in **Figure 1.5A**.

competent homodimers composed of each subunit may also form and generate off-target cleavage, which can limit safety or efficiency. To address this limitation, obligate heterodimer mutations were introduced at the dimer interface of the FokI cleavage domain which prevented homodimerization based on electrostatic and hydrophobic interactions. The creation of FokI variants that preferentially heterodimerize successfully reduced off-target cleavage of ZFNs and relieved toxicity (113-115). A similar principle was applied to TALENs by Cade and coworkers (102) for the generation of zebrafish knockout lines. The heterodimeric TALENs show similar or even greater activities than their homodimeric counterparts. Moreover, the TALENs constructed with heterodimeric FokI domains induced smaller numbers of abnormal or dead embryos, indicating reduced toxicity. This obligate heterodimeric TALEN configuration has also been reported by other groups for gene knockout studies (79,104,105).

1.4.2 DNA recognition specificity

The DNA recognition specificity of TALENs is conferred by the repeat-variable diresidues (RVDs) at positions 12 and 13 of each repeat. More than 20 different RVDs have been identified in TALEs, among which NI, NG, HD, NN, and HG are the most common ones recognizing the nucleotides A, T, C, G/A, and T, respectively (67,68). Based on crystal structures, TALE binds to target DNA as a right-handed superhelix (**Figure 1.5B**). Each repeat unit forms a left-handed, two-helix bundle that presents an RVD-containing loop to the DNA major groove (20,111). The first residue of each RVD (residue 12 of each repeat), either His or Asn, does not contact DNA directly. Instead, the side chain forms a hydrogen bond to the backbone carbonyl oxygen of Ala at position 8

(Ala^{8}) of each repeat, stabilizing the local conformation of the RVD-containing loop (**Figure 1.5 C-E**). Sequence-specific contacts of TALEs to target DNA are made by the second residue of each RVD (residue 13 of each repeat) to the corresponding base on the sense strand. In the HD RVDs specific for C nucleotides, the carboxylate oxygen of Asp^{13} forms a hydrogen bond to the amine group of cytosine, which excludes the other bases through physical or electrostatic clash (**Figure 1.5C**). In the case of NG and HG RVDs specific for T nucleotides, the backbone α carbon of the Gly^{13} makes a nonpolar van der Waals contact with the methyl group of the opposing thymine base, which is less favorable for the other bases (**Figure 1.5 D & E**). In the NI RVDs specific for A nucleotides, the aliphatic side chain of Ile^{13} makes nonpolar van der Waals interactions to C8 and N7 of adenine, which reduces the binding affinity to the other bases (**Figure 1.5F**). NN RVDs are commonly used to recognize G nucleotides. The side chain of Asn^{13} residue makes a hydrogen bond with the N7 nitrogen of the opposing guanine base (**Figure 1.5G**). But similar interaction might be made with the N7 nitrogen of adenine, which makes NN RVDs associate with A and G nucleotides with almost identical frequency. Because HD and NN RVDs form hydrogen bonds with DNA bases, the binding affinities of HD to cytosine and NN to guanine are much stronger than the van der Waals contacts of NI to adenine and NG/HG to thymine. It has been suggested to incorporate at least 3-4 strong RVDs for the construction of efficient TALENs (116).

The lack of specific RVDs to recognize guanine limits TALENs' broader applications because non-specific binding can generate off-target cleavage, resulting in unexpected genomic instability and cytotoxicity. Morbitzer and coworkers discovered that the NK RVDs can facilitate specific targeting of G nucleotides through *in planta* function

analysis (117). Based on the SELEX assay, Miller and coworkers provided *in vitro* evidence that RVD NK has a much stronger preference for guanine over adenine, which represents a promising code for the specific recognition of G nucleotides (84). However, substitution of RVD NN with NK significantly reduced TALEN activity in zebrafish embryo (105). Substantially lower activities in NK containing TALEs have also been observed in plants and in mammalian cells (116,118). Therefore, RVD NK is not ideal for guanine recognition because the improvement in specificity sacrifices efficiency. Alternatively, NH has been reported as a competent guanine-specific RVD, which has much higher efficiency than RVD NK (116,118). Computational modeling analysis showed that the imidazole ring on the His^{13} of the NH RVD has a compact base-stacking interaction with the guanine base, suggesting a possible mechanism for its increased specificity for G nucleotide while maintaining the binding affinity (118).

Although successfully used in various cellular contexts, TALE DNA binding domains have been reported to be incapable of targeting methylated DNA (119). Often considered as the fifth base, 5-methyl cytosine (5mC) is a major epigenetic mark and widely distributed in fungi, plant and mammalian genomes (120). In addition, 5mC has been identified in CpG islands of many promoters, which are important regulatory regions for genome modification (121). Recently, two groups discovered that RVD NG and N* (an asterisk indicates a deletion at residue 13 in the repeat unit) can accommodate 5mC efficiently *in vitro* and *in vivo* (112,122). Thymine is structurally similar to 5mC, with the only difference at position 4, which is not involved in binding to TALE repeats. This observation indicates that the NG RVD specific for thymine might be used to recognize 5mC. The protein crystal structure solved by Deng and coworkers shows that lack of side

chain of Gly^{13} in NG RVDs provides sufficient space to accommodate the 5-methyl group of 5mC and allows the formation of van der Waals contacts (**Figure 1.5H**) (112). Because RVDs are followed immediately by two conserved Gly residues, N* is roughly equivalent to NG except for a shortened RVD loop (**Figure 1.5I**). Using N* to code for 5mC, Valton and coworkers demonstrated the first example of TALEN-mediated modification at a methylated locus in human cells (122). Accommodation of 5mC by TALE repeats through the RVD NG or N* extends the DNA recognition code and enables researchers to design TALENs to target hypermethylated DNA regions, which has great potential in epigenetics studies and human therapeutic applications.

All naturally-occurring TALE target sites are preceded by a 5'-thymine at position 0, which was previously believed to be essential for TALE function (67,68). The TALE crystal structure reveals that two degenerate repeats prior to the central repeat domain appear to cooperate to specify the conserved 5'-thymine (111). The indole ring of a Trp residue in the repeat R(-1) forms a van der Waals contact with the methyl group of the thymine base, suggesting a possible mechanism for the conserved specificity at position 0 (**Figure 1.5J**). Sun and coworkers reported that TALENs with shorter CTSs (31 aa) show higher efficiency against natural TALE recognition sites preceded by a 5'-T than that against unnatural TALE sites preceded by A, C or G. However, TALEN variants with longer CTSs (63-117 aa) are capable of cleaving unnatural DNA substrates with similar efficiency compared with that of natural TALE sites (89). Other studies also provided evidence that a thymine at position 0 is not strictly required for TALEN activity (84,123,124). Notably, there are nine leucine zipper-like heptad repeats closely linked to the C terminus of the TALE central repeat domain (125). These leucine-rich repeats may

mediate TALE/DNA interactions and increase DNA binding affinity, making the 5'-T less of a requirement. Detailed structural studies could help solve this uncertainty. The requirement of a preceding 5'-T can be mitigated using certain TALEN scaffolds, allowing greater flexibility in choosing target sites in genome editing endeavors.

1.4.3. Assembly of TALE repeat arrays

Because of the high similarity between each TALE repeat unit, it is challenging to construct plasmids encoding long arrays of TALE repeats. To address this limitation, numerous methods have been developed to assemble the highly repetitive TALE central repeat domains rapidly and cost-effectively (64). Based on a standard cloning strategy, Sander and coworkers described a restriction enzyme and ligation (REAL) method, in which single TALE repeats are joined together using routine restriction digestion and ligation techniques (107). Initially, they constructed a library of plasmids encoding various individual TALE repeats by DNA synthesis. In the assembly step, two TALE repeats are first joined together by ligating compatible overhangs generated by digestion with restriction endonucleases. Next, the ligation product encoding two TALE repeats is joined with another TALE repeat dimer in the same manner, resulting in a DNA fragment encoding four TALE repeats. This process continues in an iterative fashion until a TALE repeat array of the desired length is assembled (**Figure 1.6A**). Using a large plasmid library of pre-assembled multiple TALE repeats, REAL can be performed in a more rapid and less labor-intensive fashion, which is referred to as REAL-Fast (126). With the help of isocaudamer restriction enzymes (e.g. *Nhe*I and *Spe*I), a unit assembly method has

been described for building long TALE repeat arrays in the similar hierarchical fashion (105).

Utilization of Golden Gate cloning has greatly facilitated and accelerated the synthesis of TALE genes (74,86,100,127-130). Golden Gate cloning has been developed to overcome the difficulty of assembling the monomers into ordered multimer arrays (131). It employs type IIS restriction endonucleases that cut outside of their recognition sequences and produce nonpalindromic, 4 bp 5'-overhangs. Since the recognition and cleavage sites are spatially separated in type IIS restriction enzymes, essentially any desired 5'-overhang sequence can be generated. Moreover, because the correct ligation products lack the enzyme recognition site and cannot be recut, the cleavage and ligation can be carried out in the same reaction mixture in a single step. To construct a TALE central repeat domain to recognize a desired target site, each individual repeat must be flanked by the type IIS restriction sites at both 5'- and 3'-ends. Restriction endonuclease digestion removes the flanking restriction sites and generates unique terminal overhangs. The overhangs are designed so that each repeat ligates specifically to another repeat with a compatible overhang. Thus, the 3'-end of first repeat can only ligate to the 5'-end of the second repeat, the 3'-end of the second repeat can only ligate to the 5'-end of the third repeat and so on. Therefore, the position of each repeat within the TALE central repeat domain is defined exclusively by the given overlap. For the first round of Golden Gate reaction, 6-10 repeats are cloned into an intermediate plasmid. The second round of Golden Gate reaction assembles the TALE repeat arrays of each intermediate plasmid into the final backbone plasmid and makes the complete TALE central repeat domain fused with FokI or other functional protein domains (**Figure 1.6B**). To increase the

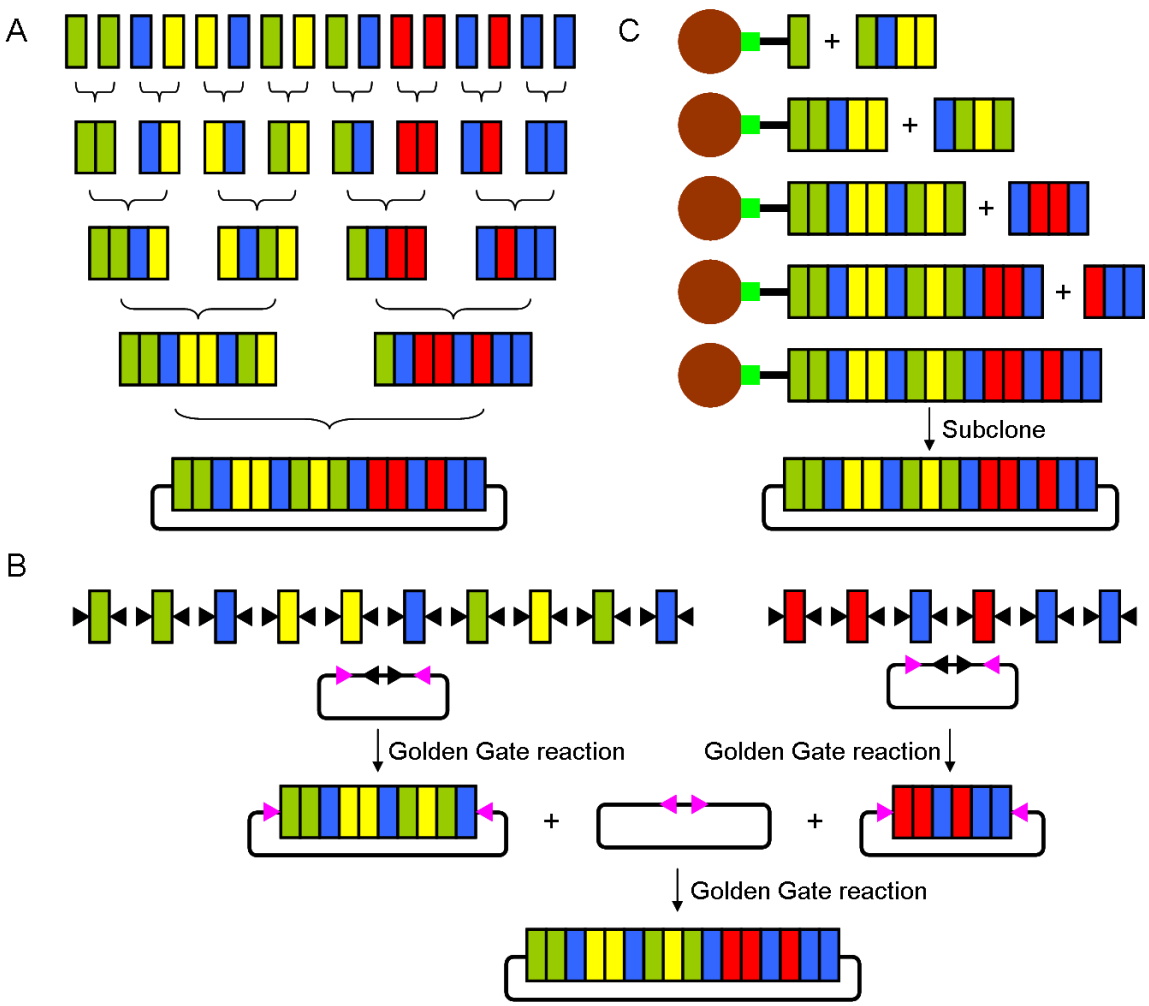

Figure 1.6. Schematic of the strategies for the assembly of TALE repeat arrays. (**A**) The REAL strategy based on hierarchical ligations (107). (**B**) The Golden Gate cloning-based strategy (74). Each repeat unit is flanked by the recognition sites of a type IIS restriction endonuclease (black triangles). The first round of Golden Gate reaction assembles multiple TALE repeat units in a single step. The second round of Golden Gate reaction relies on the recognition site of a different type IIS restriction enzyme (pink triangles), which assembles the complete TALE genes from pre-assembled repeat multimers. (**C**) The FLASH assembly method based on solid-phase ligation (85). Ligations are carried out iteratively on a streptavidin-coated magnetic bead (brown circle), which contains an immobilized biotinylated DNA double-strand adaptor (green box). The final ligation product is released from the solid phase by restriction digestion and subcloned into the final backbone vector.

assembly efficiency, the *lacZ* gene and the toxic *ccdB* gene are introduced for blue/white screening and selection, respectively. With the help of preassembled TALE repeat tetramers and trimers, a single-step Golden Gate strategy has been developed to generate TALENs that recognize 15 bp target sites within two days (69). Due to its ease of use and public availability, Golden Gate assembly provides a convenient means to construct TALENs for academic laboratories.

For industrial scale synthesis, development of a solid-phase ligation strategy has facilitated the cloning of TALE genes in a high-throughput and cost-effective manner (85,90,123). The solid-phase strategy assembles TALE repeat units on a streptavidin-coated magnetic bead that contains an immobilized biotinylated DNA double-strand adaptor with a restriction endonuclease site on one end. Ligation of TALE repeat units is unidirectional and iterative. In each cycle, newly added TALE repeat units are ligated to the immobilized DNA fragments and subsequent washing steps remove undesired products. The cycle is continued until an array of the desired length is assembled. The final ligation product is then released from the solid phase by restriction digestion and subcloned into the final backbone vector (**Figure 1.6C**). With this strategy, TALE repeat units are assembled on solid-phase rather than in solution, thereby avoiding the need for gel isolation, purification or analysis of intermediate plasmids. With 376 archived plasmids encoding TALE repeats as tetramers, trimers, dimers and monomers, the fast ligation-based automatable solid-phase high-throughput (FLASH) system enables assembly of 96 TALE genes in less than one day. With automation, FLASH can make sequence-verified TALE expression plasmids for <$100 each, including the cost of labor (85). Instead of using a pre-assembled plasmid library, iterative capped assembly (ICA)

builds full-length arrays from individual TALE repeat monomers (123). Introduction of capping oligonucleotides eliminates incomplete ligation and monomer self-ligation, which are essential for the production of pure full-length TALE repeat arrays. With automation, ICA enables efficient assembly of TALE genes bearing up to 21 repeats followed by ligation into an expression plasmid within three hours. Wang and coworkers performed solid-phase ligation on a chip, which allows the synthesis of >100 TALE genes bearing 16 or 20 repeats in three days (90).

Recently, a ligation-independent cloning (LIC) technique has been developed for high-throughput assembly of TALE genes (87). Compared with Golden Gate cloning, LIC relies on much longer (10-30 bp) nonpalindromic overhangs to anneal with the overhangs of other fragments in a highly specific manner. The long overhangs are generated by the controllable 3'-exonuclease activity of T4 DNA polymerase. Because the fragments' long overlaps do not dissociate during transformation, the annealed products can be directly transformed into *Escherichia coli* without prior ligation step and ligated through bacterial ligases. Because of its high fidelity, LIC circumvents agar-based single-colony picking step, which allows growth of cells directly in polyclonal cultures after transformation. Using 64 repeat dimer-containing plasmids, LIC allows generation of correctly assembled TALE genes bearing 18.5 repeat units in three days through a hierarchical, two step assembly process. In addition, a comprehensive 5-mer TALE repeat unit fragment library composed of 3072 plasmids was created, which enables automated assembly of >600 TALE genes bearing 15.5 repeat units in one day.

1.4.4 Future perspectives

The last three years witnessed the tremendous progress of the TALEN technology. The scaffold optimization isolated TALEN variants with high DNA cleavage efficiency, which is essential for targeted genome editing. The characterization of novel RVDs extended the DNA recognition code and helped to minimize off-target cleavage activity of TALENs by increasing guanine recognition specificity. Development of novel strategies for convenient and quick assembly of TALE repeat arrays enabled high-throughput synthesis of TALENs and made TALEN technology accessible and affordable for any academic or industrial lab.

Besides TALENs, there are other tools available for editing genomes. Meganucleases are natural DNA endnucleases with high activity and specificity, but it is difficult to tailor their DNA recognition specificities. It is relatively easier to engineer ZFNs to target custom-designed DNA sequences, but some of them suffer from requirement of intensive labor for construction and off-target effects (as reviewed in (132)). Recently, Clustered regularly interspaced short palindromic repeats (CRISPR)-mediated DNA cleavage has been applied for genome editing (133-138). This system can be reprogrammed readily using customized RNAs and enable multiplex genome engineering. However, the limited target specificity (14 bp) can cause off-target cleavage and the requirement for a protospacer adjacent motif (PAM) restricts its targeting range. Compared with these tools, TALENs have the advantage of high specificity and modularity, but there are limitations that remain to be addressed for their further improvement. The bulky size of TALENs might limit their broader applications, especially in the cases when efficient gene delivery cannot be achieved. Development of strategies for efficient delivery of TALEN genes

into cells would enable TALEN-mediated genome editing in more different organisms and cell types. In eukaryotic cells, DNA is packaged into chromatin. Therefore, chromosomal context and epigenetic modifications play a major role in the DNA accessibility of TALENs. The combination of epigenetic modification tools and TALEN technology could expand the range of target for TALEN-mediated genome modifications, which might be a potential area for future exploration. Unlike ZFNs, the off-target effects of TALENs have not been comprehensively characterized. Mussolino and coworkers carried out a side-by-side comparison between ZFNs and TALENs and found significantly reduced nuclease-associated cytotoxicities of TALENs (70). Ding and coworkers also reported minimal off-target-effects of TALENs using exome sequencing whole-genome sequencing at low coverage, but they still could not completely rule out TALEN off-target effects (69). Therefore, careful screening of the complete genome of TALEN-modified cells using deep sequencing analysis would be instructive for safe use of TALENs, especially for human clinical applications.

Other than fusing with FokI to make DNA endonucleases, TALEs have been used to create novel chimeric proteins by fusing with other functional protein domains. TALE-based transcription activators have been constructed to induce transcription of endogenous genes in plants (117) and human cells (74,119,127,139-141). By fusing with transcription repressor domains, TALEs have been used to generate artificial repressors for sequence-specific gene repression in bacteria (142), yeast (143), plants (144) and human cells (118,141). In addition, chimeric TALE recombinases (TALERs) have been constructed by fusing a hyperactivated catalytic domain from the DNA invertase Gin with a TALE central repeat domain (145). TALERs with optimized architecture

recombine DNA efficiently in bacterial and mammalian cells, providing an alternative approach for targeted genome editing. It would be interesting to combine TALE central repeat arrays with other different functional domains for many different applications in the future. For example, a TALE combined with a ligand-binding domain can be applied for high-throughput drug screening; a TALE combined with a DNA methyltransferase can be used for targeted DNA modification; a TALE combined with a histone deacetylase can be used for specific chromatin modification; a TALE combined with a cytosine deaminase can be applied for endogenous targeted mutagenesis, etc.

Thanks to simplicity in design, convenience in construction and high success rates across species, TALEN technology has received much attention since its invention. Although challenges and obstacles remain, TALEN technology will continue to be an important topic for future research and development and benefit both basic and applied biological sciences.

1.5 Project overview

My thesis research focuses on the design, construction, optimization and application of TALENs as an efficient tool for genome engineering. Because the wild-type TALEs are transcription activators for bacterial infection of host plants, their sequences other than the DNA binding domain can be detrimental to the nuclease activity once fused with the FokI DNA cleavage domain. To maximize TALEN efficiency, protein engineering strategies including scaffold optimization and directed evolution have been utilized in yeast cells. A high-throughput screening of a linker library identified a novel TALEN architecture that can cleave the DNA target as a monomer. To demonstrate their

applications in human therapy, I applied the optimized TALENs for use in the treatment of sickle cell disease. TALEN-mediated genome editing enabled the correction of the disease-causing gene in induced pluripotent stem cells isolated from a patient.

Chapter 2 describes scaffold optimization of TALENs for use in treatment of sickle cell disease. By using a yeast reporter system, a systematic study was carried out to optimize TALEN architecture for maximal *in vivo* cleavage efficiency. In contrast to the previous reports, the engineered TALENs were capable of recognizing and cleaving target binding sites preceded by A, C or G. More importantly, the optimized TALENs efficiently cleaved a target sequence within the human β-globin (*HBB*) gene associated with sickle cell disease and increased the efficiency of targeted gene repair by >1000-fold in human cells. In addition, these TALENs showed no detectable cytotoxicity. These results demonstrate the potential of optimized TALENs as a powerful genome editing tool for therapeutic applications.

Chapter 3 reports the seamless gene correction of the sickle cell disease mutation in human induced pluripotent stem cells using TALENs. The TALENs I have engineered are highly specific and generate minimal off-target effects. In combination with *piggyBac* transposon, TALEN-mediated gene targeting leaves no residual ectopic sequences at the site of correction and the corrected stem cells retain full pluripotency and a normal karyotype. This study demonstrates an important first step of using TALENs for the treatment of genetic diseases, which represents a significant advance toward cell and gene therapies.

Chapter 4 reports the development of a TALEN variant, SunnyTALEN, with >2.5-fold improved genome editing efficacy in human cells. A high-throughput screening

system has been constructed in yeast cells in order to improve TALEN efficiency through directed evolution. After multiple rounds of mutagenesis and screening, 14 TALEN mutants were isolated with improved activity either in yeast or human cells, among which SunnyTALEN shows highest activity for human genome editing. The corresponding scaffold increases the rate of genetic modification at all the 13 tested loci of human genome and is compatible with heterodimer TALEN architectures. This enhanced and high-efficiency TALEN variant represents a novel second-generation TALEN system and has great potential for biological and therapeutic applications.

Finally, Chapter 5 provides preliminary results of the development and characterization of a novel TALEN architecture, single-chain TALEN (scTALEN). Compared with the conventional TALEN scaffold, scTALEN has two FokI DNA cleavage domains which are linked by a polypeptide linker. The appropriate linker was isolated from a polypeptide linker library from high-throughput screening. Instead of forming a heterodimer, the scTALEN we identified can cleave the DNA substrate as a monomer, with the formation of a catalytic active intra-molecular FokI dimer. The corresponding scTALEN has moderate *in vivo* activity in both yeast and human cells. Compared with the conventional TALEN architecture, scTALEN has only half the protein size, which significantly decreases the gene delivery payload, especially for certain cell types such as primary human cells with low DNA or RNA delivery efficiency.

1.6 References

1. Levasseur, D.N., Ryan, T.M., Pawlik, K.M. and Townes, T.M. (2003) Correction of a mouse model of sickle cell disease: lentiviral/antisickling beta-globin gene

transduction of unmobilized, purified hematopoietic stem cells. *Blood*, **102**, 4312-4319.

2. Check, E. (2002) A tragic setback. *Nature*, **420**, 116-118.
3. Marshall, E. (1999) Gene therapy death prompts review of adenovirus vector. *Science*, **286**, 2244-2245.
4. Gersbach, C.A., Gaj, T., Gordley, R.M., Mercer, A.C. and Barbas, C.F., 3rd. (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic Acids Res*, **39**, 7868-7878.
5. Feng, X., Bednarz, A.L. and Colloms, S.D. (2010) Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. *Nucleic Acids Res*, **38**, 1204-1216.
6. Jensen, N.M., Dalsgaard, T., Jakobsen, M., Nielsen, R.R., Sorensen, C.B., Bolund, L. and Jensen, T.G. (2011) An update on targeted gene repair in mammalian cells: methods and mechanisms. *J Biomed Sci*, **18**, 10.
7. Arnould, S., Delenda, C., Grizot, S., Desseaux, C., Paques, F., Silva, G.H. and Smith, J. (2010) The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. *Protein Eng Des Sel*, **24**, 27-31.
8. Marcaida, M.J., Munoz, I.G., Blanco, F.J., Prieto, J. and Montoya, G. (2010) Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci*, **67**, 727-748.

9. Chevalier, B.S. and Stoddard, B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res,* **29**, 3757-3774.

10. Chames, P., Epinat, J.C., Guillier, S., Patin, A., Lacroix, E. and Paques, F. (2005) In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res*, **33**, e178.

11. Chen, Z. and Zhao, H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res,* **33**, e154.

12. Doyon, J.B., Pattanayak, V., Meyer, C.B. and Liu, D.R. (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc,* **128**, 2477-2484.

13. Chen, Z., Wen, F., Sun, N. and Zhao, H. (2009) Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Eng Des Sel*, **22**, 249-256.

14. Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J., Jr., Stoddard, B.L. and Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656-659.

15. Ashworth, J., Taylor, G.K., Havranek, J.J., Quadri, S.A., Stoddard, B.L. and Baker, D. (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res*, **38**, 5601-5608.

16. Thyme, S.B., Jarjour, J., Takeuchi, R., Havranek, J.J., Ashworth, J., Scharenberg, A.M., Stoddard, B.L. and Baker, D. (2009) Exploitation of binding energy for catalysis and design. *Nature*, **461**, 1300-1304.

17. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S. *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol*, **355**, 443-458.

18. Redondo, P., Prieto, J., Munoz, I.G., Alibes, A., Stricher, F., Serrano, L., Cabaniols, J.P., Daboussi, F., Arnould, S., Perez, C. *et al.* (2008) Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature*, **456**, 107-111.

19. Kim, C.A. and Berg, J.M. (1996) A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat Struct Biol*, **3**, 940-945.

20. Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y. and Yan, N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720-723.

21. Fajardo-Sanchez, E., Stricher, F., Paques, F., Isalan, M. and Serrano, L. (2008) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res*, **36**, 2163-2173.

22. Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*, **31**, 2952-2962.

23. Gao, H., Smith, J., Yang, M., Jones, S., Djukanovic, V., Nicholson, M.G., West, A., Bidney, D., Falco, S.C., Jantz, D. *et al.* (2010) Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J*, **61**, 176-187.

24. Li, H., Pellenz, S., Ulge, U., Stoddard, B.L. and Monnat, R.J., Jr. (2009) Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res*, **37**, 1650-1662.

25. Ulge, U.Y., Baker, D.A. and Monnat, R.J., Jr. (2011) Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res*, **39**, 4330-4339.

26. Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J. *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res*, **34**, e149.

27. Munoz, I.G., Prieto, J., Subramanian, S., Coloma, J., Redondo, P., Villate, M., Merino, N., Marenchino, M., D'Abramo, M., Gervasio, F.L. *et al.* (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res*, **39**, 729-743.

28. Arnould, S., Perez, C., Cabaniols, J.P., Smith, J., Gouble, A., Grizot, S., Epinat, J.C., Duclert, A., Duchateau, P. and Paques, F. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol*, **371**, 49-65.

29. Grosse, S., Huot, N., Mahiet, C., Arnould, S., Barradeau, S., Clerre, D.L., Chion-Sotinel, I., Jacqmarcq, C., Chapellier, B., Ergani, A. *et al.* (2011) Meganuclease-mediated inhibition of HSV1 infection in cultured cells. *Mol Ther*, **19**, 694-702.

30. Grizot, S., Smith, J., Daboussi, F., Prieto, J., Redondo, P., Merino, N., Villate, M., Thomas, S., Lemaire, L., Montoya, G. *et al.* (2009) Efficient targeting of a SCID

gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res*, **37**, 5405-5419.

31. Takeuchi, R., Lambert, A.R., Mak, A.N., Jacoby, K., Dickson, R.J., Gloor, G.B., Scharenberg, A.M., Edgell, D.R. and Stoddard, B.L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci U S A*, **108**, 13077-13082.

32. Porteus, M.H. and Carroll, D. (2005) Gene targeting using zinc finger nucleases. *Nat Biotechnol*, **23**, 967-973.

33. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10570-10575.

34. Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773-782.

35. Cathomen, T. and Joung, J.K. (2008) Zinc-finger nucleases: the next generation emerges. *Mol Ther*, **16**, 1200-1207.

36. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, **11**, 636-646.

37. Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol*, **29**, 816-823.

38. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods*, **8**, 765-770.

39. Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol*, **26**, 808-816.

40. Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C. *et al.* (2010) Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat Biotechnol*, **28**, 839-847.

41. Cradick, T.J., Keck, K., Bradshaw, S., Jamieson, A.C. and McCaffrey, A.P. (2010) Zinc-finger nucleases as a novel therapeutic strategy for targeting hepatitis B virus DNAs. *Mol Ther*, **18**, 947-954.

42. Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., Jenkins, S.S., Wood, A., Cui, X., Meng, X. *et al.* (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*, **325**, 433.

43. Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T. and Serikawa, T. (2010) Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS One*, **5**, e8870.

44. Flisikowska, T., Thorey, I.S., Offner, S., Ros, F., Lifke, V., Zeitler, B., Rottmann, O., Vincent, A., Zhang, L., Jenkins, S. *et al.* (2011) Efficient immunoglobulin

gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PLoS One*, **6**, e21045.

45. Hauschild, J., Petersen, B., Santiago, Y., Queisser, A.L., Carnwath, J.W., Lucas-Hahn, A., Zhang, L., Meng, X., Gregory, P.D., Schwinzer, R. *et al.* (2011) Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc Natl Acad Sci U S A*, **108**, 12013-12017.

46. Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q. *et al.* (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*, **5**, 97-110.

47. Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*, **27**, 851-857.

48. Kotin, R.M., Linden, R.M. and Berns, K.I. (1992) Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *Embo J*, **11**, 5071-5078.

49. DeKelver, R.C., Choi, V.M., Moehle, E.A., Paschon, D.E., Hockemeyer, D., Meijsing, S.H., Sancak, Y., Cui, X., Steine, E.J., Miller, J.C. *et al.* (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res*, **20**, 1133-1142.

50. Zou, J., Sweeney, C.L., Chou, B.K., Choi, U., Pan, J., Wang, H., Dowey, S.N., Cheng, L. and Malech, H.L. (2011) Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood*, **117**, 5561-5572.

51. Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F. *et al.* (2011) In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*, **29**, 1717-1726.

52. Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. *et al.* (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell*, **31**, 294-301.

53. Zou, J., Mali, P., Huang, X., Dowey, S.N. and Cheng, L. (2011) Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood*, **118**, 4599-4608.

54. Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordonez, A., Hannan, N.R., Rouhani, F.J. *et al.* (2011) Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, **478**, 391-394.

55. Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A. and Liu, P. (2008) Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci U S A*, **105**, 9290-9295.

56. Soldner, F., Laganiere, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, **146**, 318-331.

57. Feuk, L. (2010) Inversion variants in the human genome: role in disease and genome architecture. *Genome medicine*, **2**, 11.

58. Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M. and Jasin, M. (2009) Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A*, **106**, 10620-10625.

59. Lee, H.J., Kim, E. and Kim, J.S. (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res*, **20**, 81-89.

60. Lee, H.J., Kweon, J., Kim, E., Kim, S. and Kim, J.S. (2012) Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res*, **22**, 539-548.

61. Bagnall, R.D., Waseem, N., Green, P.M. and Giannelli, F. (2002) Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood*, **99**, 168-174.

62. Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods*, **5**, 374-375.

63. Radecke, S., Radecke, F., Cathomen, T. and Schwarz, K. (2010) Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol Ther*, **18**, 743-753.

64. Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, **14**, 49-55.

65. Bogdanove, A.J. and Voytas, D.F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science*, **333**, 1843-1846.

66. Munoz Bodnar, A., Bernal, A., Szurek, B. and Lopez, C.E. (2013) Tell me a tale of TALEs. *Mol Biotechnol*, **53**, 228-235.

67. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.

68. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.

69. Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T. *et al.* (2013) A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. *Cell Stem Cell*, **12**, 238-251.

70. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, **39**, 9283-9293.

71. Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res*, **39**, 359-372.

72. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757-761.

73. White, F.F., Potnis, N., Jones, J.B. and Koebnik, R. (2009) The type III effectors of Xanthomonas. *Mol Plant Pathol*, **10**, 749-766.

74. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res*, **39**, e82.

75. Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D.F., Zheng, X. *et al.* (2013) Rapid and efficient gene modification in rice and Brachypodium using TALENs. *Mol Plant*, doi: 10.1093/mp/sss1162.

76. Carlson, D.F., Tan, W., Lillico, S.G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D.F., Long, C.R., Whitelaw, C.B. and Fahrenkrug, S.C. (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A*, **109**, 17382-17387.

77. Watanabe, T., Ochiai, H., Sakuma, T., Horch, H.W., Hamaguchi, N., Nakamura, T., Bando, T., Ohuchi, H., Yamamoto, T., Noji, S. *et al.* (2012) Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat Commun*, **3**, 1017.

78. Ishibashi, S., Cliffe, R. and Amaya, E. (2012) Highly efficient bi-allelic mutation rates using TALENs in Xenopus tropicalis. *Biol Open*, **1**, 1273-1276.

79. Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., Cheng, C.H., Dawid, I.B., Chen, Y. and Zhao, H. (2012) Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc Natl Acad Sci U S A*, **109**, 17484-17489.

80. Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B. *et al.* (2012) Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *J Genet Genomics*, **39**, 209-215.

81. Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M.C., Urnov, F.D., Gregory, P.D. and Cost, G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol Bioeng*, **110**, 871-880.

82. Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C. *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*, **29**, 731-734.

83. Kim, H., Um, E., Cho, S.R., Jung, C., Kim, H. and Kim, J.S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods*, **8**, 941-943.

84. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, **29**, 143-148.

85. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, **30**, 460-465.

86. Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G. and Zhang, F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc*, **7**, 171-192.

87. Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol*, **31**, 76-81.

88. Stroud, D.A., Formosa, L.E., Wijeyeratne, X.W., Nguyen, T.N. and Ryan, M.T. (2013) Gene knockout using transcription activator-like effector nucleases (TALENs) reveals that human NDUFA9 protein is essential for stabilizing the junction between membrane and matrix arms of complex I. *J Biol Chem*, **288**, 1685-1690.

89. Sun, N., Liang, J., Abil, Z. and Zhao, H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst*, **8**, 1255-1263.

90. Wang, Z., Li, J., Huang, H., Wang, G., Jiang, M., Yin, S., Sun, C., Zhang, H., Zhuang, F. and Xi, J.J. (2012) An integrated chip for the high-throughput synthesis of transcription activator-like effectors. *Angew Chem Int Ed Engl*, **51**, 8505-8508.

91. Ansai, S., Sakuma, T., Yamamoto, T., Ariga, H., Uemura, N., Takahashi, R. and Kinoshita, M. (2013) Efficient targeted mutagenesis in medaka using custom-designed transcription activator-like effector nucleases. *Genetics*, **193**, 739-749.

92. Sung, Y.H., Baek, I.J., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.S. and Lee, H.W. (2013) Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol*, **31**, 23-24.

93. Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.

94. Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X., Zhang, L. *et al.* (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol*, **29**, 695-696.

95. Tong, C., Huang, G., Ashton, C., Wu, H., Yan, H. and Ying, Q.L. (2012) Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs. *J Genet Genomics*, **39**, 275-280.

96. Li, T., Liu, B., Spalding, M.H., Weeks, D.P. and Yang, B. (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol*, **30**, 390-392.

97. Ma, S., Zhang, S., Wang, F., Liu, Y., Liu, Y., Xu, H., Liu, C., Lin, Y., Zhao, P. and Xia, Q. (2012) Highly efficient and specific genome editing in silkworm using custom TALENs. *PLoS One*, **7**, e45035.

98. Sajwan, S., Takasu, Y., Tamura, T., Uchino, K., Sezutsu, H. and Zurovec, M. (2013) Efficient disruption of endogenous Bombyx gene by TAL effector nucleases. *Insect Biochem Mol Biol*, **43**, 17-23.

99. Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J. and Voytas, D.F. (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol*, **161**, 20-27.

100. Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*, **39**, 6315-6325.

101. Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y. *et al.* (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114-118.

102. Cade, L., Reyon, D., Hwang, W.Y., Tsai, S.Q., Patel, S., Khayter, C., Joung, J.K., Sander, J.D., Peterson, R.T. and Yeh, J.R. (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res*, **40**, 8001-8010.

103. Chen, S., Oikonomou, G., Chiu, C.N., Niles, B.J., Liu, J., Lee, D.A., Antoshechkin, I. and Prober, D.A. (2013) A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res*, **41**, 2769-2778.

104. Dahlem, T.J., Hoshijima, K., Jurynec, M.J., Gunther, D., Starker, C.G., Locke, A.S., Weis, A.M., Voytas, D.F. and Grunwald, D.J. (2012) Simple methods for

generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet*, **8**, e1002861.

105. Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S. and Zhang, B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol*, **29**, 699-700.

106. Moore, F.E., Reyon, D., Sander, J.D., Martinez, S.A., Blackburn, J.S., Khayter, C., Ramirez, C.L., Joung, J.K. and Langenau, D.M. (2012) Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PLoS One*, **7**, e37877.

107. Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K. and Yeh, J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol*, **29**, 697-698.

108. Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A*, **108**, 2623-2628.

109. Christian, M.L., Demorest, Z.L., Starker, C.G., Osborn, M.J., Nyquist, M.D., Zhang, Y., Carlson, D.F., Bradley, P., Bogdanove, A.J. and Voytas, D.F. (2012) Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS One*, **7**, e45383.

110. Gao, H., Wu, X., Chai, J. and Han, Z. (2012) Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res*, **22**, 1716-1720.

111. Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. and Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716-719.

112. Deng, D., Yin, P., Yan, C., Pan, X., Gong, X., Qi, S., Xie, T., Mahfouz, M., Zhu, J.K., Yan, N. *et al.* (2012) Recognition of methylated DNA by TAL effectors. *Cell Res*, **22**, 1502-1504.

113. Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol*, **25**, 786-793.

114. Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*, **25**, 778-785.

115. Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D. and Holmes, M.C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods*, **8**, 74-79.

116. Streubel, J., Blucher, C., Landgraf, A. and Boch, J. (2012) TAL effector RVD specificities and efficiencies. *Nat Biotechnol*, **30**, 593-595.

117. Morbitzer, R., Romer, P., Boch, J. and Lahaye, T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci U S A*, **107**, 21617-21622.

118. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. and Zhang, F. (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun*, **3**, 968.

119. Bultmann, S., Morbitzer, R., Schmidt, C.S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T. and Leonhardt, H. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res*, **40**, 5368-5377.

120. Su, Z., Han, L. and Zhao, Z. (2011) Conservation and divergence of DNA methylation in eukaryotes: new insights from single base-resolution DNA methylomes. *Epigenetics*, **6**, 134-140.

121. Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y. *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, **466**, 253-257.

122. Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliand, K., Juillerat, A. and Duchateau, P. (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem*, **287**, 38427-38432.

123. Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P. and Church, G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res*, **40**, e117.

124. Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., Feng, J., Verdier, V. and Szurek, B. (2011) Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae pv. oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Mol Plant Microbe Interact*, **24**, 1102-1113.

125. Yang, Y. and Gabriel, D.W. (1995) Xanthomonas avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. *Mol Plant Microbe Interact*, **8**, 627-631.

126. Reyon, D., Khayter, C., Regan, M.R., Joung, J.K. and Sander, J.D. (2012) Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Curr Protoc Mol Biol*, **Chapter 12**, Unit 12 15.

127. Geissler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.E. and Boch, J. (2011) Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*, **6**, e19509.

128. Li, L., Piatek, M.J., Atef, A., Piatek, A., Wibowo, A., Fang, X., Sabir, J.S., Zhu, J.K. and Mahfouz, M.M. (2012) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol Biol*, **78**, 407-416.

129. Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res*, **39**, 5790-5799.

130. Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, **6**, e19722.

131. Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**, e3647.

132. Sun, N., Abil, Z. and Zhao, H. (2012) Recent advances in targeted genome engineering in mammalian systems. *Biotechnol J*, **7**, 1074-1087.

133. Cho, S.W., Kim, S., Kim, J.M. and Kim, J.S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*, **31**, 230-232.

134. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

135. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R. and Joung, J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*, **31**, 227-229.

136. Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, **31**, 233-239.

137. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.

138. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.

139. Garg, A., Lohmueller, J.J., Silver, P.A. and Armel, T.Z. (2012) Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res*, **40**, 7584-7595.

140. Tremblay, J.P., Chapdelaine, P., Coulombe, Z. and Rousseau, J. (2012) Transcription activator-like effector proteins induce the expression of the frataxin gene. *Hum Gene Ther*, **23**, 883-890.

141. Li, Y., Moore, R., Guinn, M. and Bleris, L. (2012) Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci Rep*, **2**, 897.

142. Politz, M.C., Copeland, M.F. and Pfleger, B.F. (2013) Artificial repressors for controlling gene expression in bacteria. *Chem Commun (Camb)*, **49**, 4325-4327.

143. Blount, B.A., Weenink, T., Vasylechko, S. and Ellis, T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One*, **7**, e33279.

144. Mahfouz, M.M., Li, L., Piatek, M., Fang, X., Mansour, H., Bangarusamy, D.K. and Zhu, J.K. (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol*, **78**, 311-321.

145. Mercer, A.C., Gaj, T., Fuller, R.P. and Barbas, C.F., 3rd. (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res*, **40**, 11163-11172.

CHAPTER 2. SCAFFOLD OPTIMIZATION OF TALENS FOR USE IN TREATMENT OF SICKLE CELL DISEASE

2.1. Introduction

Sickle cell disease (SCD) is one of the most prevalent autosomal recessive disorders worldwide, with which approximately 250,000 infants are born each year (1). The homozygous mutation causing SCD is a single substitution (from A to T) in the codon for amino acid 6 on the human β-globin (*HBB*) gene. The replacement of glutamate by valine causes the body to produce an abnormal hemoglobin protein that tends to aggregate, making the red blood cells to become stiff and develop a sickle form. These sickle red blood cells can block blood vessels, thus reducing blood flow to many parts of the body that results in tissue and organ damage (reviewed by (2)). *Ex vivo* gene therapy by gene addition has proven to be an effective method to cure SCD in mouse models (3,4). However, these studies utilize viral vectors that could lead to a toxic immunological response (5). In addition, random insertion of foreign DNA fragments into human chromosomes may result in disruption of essential genes, thus causing unpredictable side effects (e.g., carcinogenesis) (6).

Targeted gene correction is considered a promising strategy to cure SCD with minimal side effects because it can directly correct the mutation on the chromosome, thus preserving temporal and tissue-specific expression of the impaired *HBB* gene (7). Targeted gene correction involves the homologous recombination (HR) of chromosomal DNA sequences with exogenous DNA sequences to provide a precise method to systematically edit mammalian genomes (8). The rates of recombination in mammalian

cells typically fall in the range of 10^{-6}–10^{-5} events/cell/generation (9), which is an inherent limitation of this technique. However, chromosomal double strand breaks (DSBs) at a desired chromosomal locus can stimulate HR by upwards of 1000 fold to facilitate gene targeting (see **Chapter 1**).

Zinc finger nucleases (ZFNs) show promise in improving the efficiency of gene targeting by generating DSBs at preselected sites on the chromosome (10,11). ZFNs are artificial DNA nucleases constructed by fusing the custom-designed zinc finger domain with the non-specific DNA cleavage domain of the FokI type IIS restriction endonucleases. Although the ZFN technology has been successfully applied to genome editing in many organisms including human cell lines (12), generating a sequence-specific ZFN with novel specificity and high affinity remains an overwhelming challenge (13,14). Moreover, the lack of specificity of the zinc finger domains may lead to recognition of off-target sites, causing undesired genomic instability (15), which hampers its application in gene targeting based gene therapy.

Transcription activator-like (TAL) effector nucleases (TALENs) are a new class of artificial nucleases that recognize long, specific DNA sequences (as reviewed in (16)). TALENs utilize the central repeat domain of TAL effectors (TALEs) as the DNA recognition module and the FokI catalytic domain as the DNA cleavage module. TALEs bind their target sites through a central repeat domain consisting of a varying number of repeat units of 33-35 amino acids. Each repeat is largely identical except for two highly variable amino acids at positions 12 and 13, referred to as the repeat variable di-residues (RVDs). Each repeat recognizes a single nucleotide, and the recognition specificity is determined by the RVDs (e.g., NI recognizes A, HD recognizes C, NG or HG recognizes

T, and NN recognizes G or A). The DNA recognition code provides a one-to-one correspondence between the array of amino acid repeats and the nucleotide sequence of the DNA target (17). This simple DNA recognition code and its modular nature make TALEs an ideal platform for constructing custom-designed artificial DNA nucleases.

In this chapter, we sought to optimize TALEN scaffolds for use in treatment of SCD. By using a yeast based reporter system, we maximized the *in vivo* cleavage efficiency of TALENs by varying the N- and C-terminal extensions flanking the central repeat domain and the spacer length between each effector binding element (EBE) of the DNA substrates. In addition, we discovered that some TALEN architectures are capable of recognizing "unnatural" EBEs not preceded by a 5'-T. We constructed a pair of custom-designed TALENs to recognize a target sequence within the *HBB* gene locus. This pair of TALENs with the optimized configurations allowed for efficient cleavage at the desired site and enhanced the HR rate by >1000-fold with no detectable cytotoxicity. This work describes the first example of engineered TALENs capable of recognizing and cleaving a human disease-associated gene. Such TALENs represent a promising approach to correct disease-causing point mutations *in situ* and serve as a powerful genome editing tool in gene therapy.

2.2. Results

2.2.1. Construction and optimization of TALENs

AvrXa10 is a TALE from *Xanthomonas oryzae* pv. *oryzae* (18) containing a central region of 15.5 repeat units, with the last half repeat containing only the first 20 residues of other repeats. All of the reported TALEs require the first nucleotide of the binding site

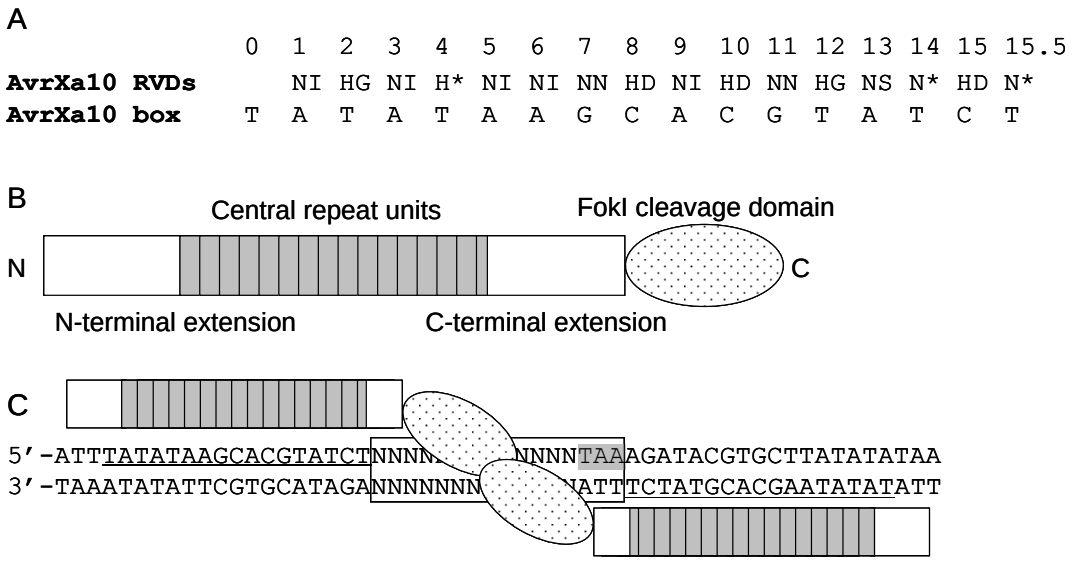

Figure 2.1. Structure and sequence of TALEN. (**A**) RVDs of the 15.5 repeat units of AvrXa10 with its DNA target sequence shown below. An asterisk indicates a deletion at residue 13. (**B**) Schematic of TALEN. N, N-terminus; C, C-terminus. (**C**) Palindromic target sites (PTSs) of TALEN-AvrXa10 are composed of two AvrXa10 EBEs (underlined) in an inverted orientation on different DNA strands, separated by a spacer (boxed) harboring an in-frame TAA stop codon (shaded in grey). TALEN-AvrXa10 functions as a homodimer.

to be preceded by a 5'-T at position 0 (17), meaning that the 15.5 repeats recognize a 17 bp EBE (**Figure 2.1A**). We created an AvrXa10 based TALEN by fusing the central repeat domain of AvrXa10 with the FokI cleavage domain (**Figure 2.1B**). Since FokI dimerization is required for efficient DNA cleavage (19), two AvrXa10 EBE sites were placed in a tail-to-tail orientation separated by a spacer of varying lengths (6-32 bp), herein referred to as palindromic target sites (PTSs, **Figure 2.1C**).

To determine the *in vivo* cleavage efficiency of a TALEN, we established a yeast reporter system adapted from a previously reported single strand annealing assay (20). The reporter plasmid contains a divided *lacZ* gene in which a duplicated 100 bp portion

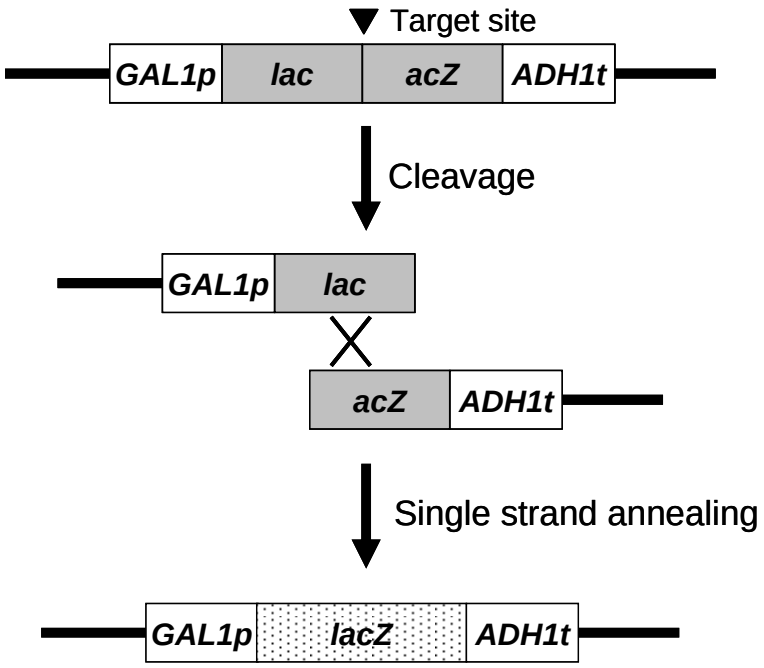

Figure 2.2. Schematic of the yeast reporter system. Details are provided in the text. Recovered *lacZ* gene was shown in dotted box.

of the LacZ coding region has been created. The two truncated *lacZ* DNA fragments are separated by a PTS containing an in-frame stop codon between the two EBE sites. Once the DSB between the two *lacZ* fragments is executed, the 100 bp direct DNA repeats will undergo HR and thus reconstitute a complete and functional *lacZ* gene (**Figure 2.2**). The β-GAL activity was used to quantify the HR rate, which correlates to the *in vivo* activity of the TALENs. Homing endonuclease I-CreI served as an internal control.

The DNA fragment of the AvrXa10-encoding gene digested by *Stu*I and *Aat*II encompasses the coding sequence for the central repeat domain, 50 amino acids upstream, and 31 amino acids downstream, with the FokI catalytic domain fused to create a minimal TALEN (referred to as N0-C0, **Figure 2.3**). No activity was observed for N0-C0 targeting any given DNA substrate (**Figure 2.4** and **Table 2.1**). Since the N-terminal

```
MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPA
▲N4(+287)
PSPAFSAGSFSDLLRQFDPSLLDTSLLDSMPAVGTPHTAAAPAECDEVQSGLRAADDPP
                    ▲N3(+207)
PTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVA
                                      ▲N2(+135) ▲N1(+125)
QHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGAR
ALEALLTEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLN
  ▲N0(+50)

------------------- Central Repeat Domain -----------------

LTPDQVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGL
                                                    ▲C0(+31)
PHAPELIRRINRRIPERTSHRVADLAHVVRVLGFFQSHSHPAQAFDDAMTQFGMSRHGL
                      ▲C1(+63)
AQLFRRVGVTELEARYGTLPPASQRWDRILQASGMKRAKPSPTSAQTPDQASLHAYYKD
                  ▲C2(+117)
DDDKKGRPSPMHEGDQTRASSRKRSRSDRAVTGPSTQQSFEVRVPEQQDALHLPLSWRV
    ▲C3(+163)                               ▲C4(+200)
KRPRTRIGGGLPDPGTPIAADLAASSTVMWEQDAAPFAGAADDFPAFNEEELAWLMELL
PQSGSVGGTI
         ▲C5(+286)
```

Figure 2.3. The N- and C-terminal amino acid sequences of TALE AvrXa10. Positions of N- and C-terminal extensions of all TALEN variants are shown and the values indicate the distance from the N- or C-terminus to the central repeat units.

sequence of a TALE is important for its DNA binding (21), we set out to add additional N-terminal extensions from AvrXa10 to the N0-C0 scaffold to achieve DNA cleavage with high efficiency. The N1-C0 construct bearing a 125 amino acid N-terminal extension showed robust nuclease activity against PTSs with 10-16 bp spacers, with the cleavage efficiency over 2-fold compared to I-CreI. The N2-C0 and N3-C0 variants with longer N-terminal extensions remained active in cleaving PTSs with 10-16 bp spacers, and showed increased nuclease activity for longer DNA substrates, such as PTSs with a 24 bp or a 32 bp spacer. The N4-C0 construct harboring the full length N-terminus from vrXa10 did not further increase activity toward PTSs with 10-16 bp spacers but showed

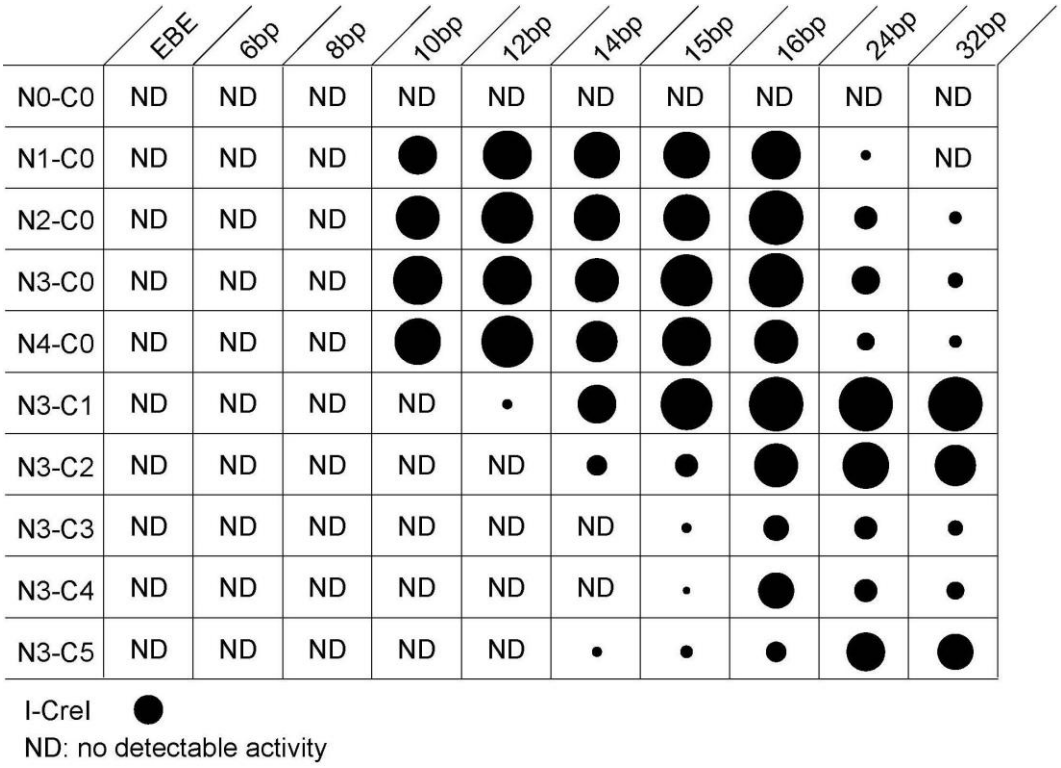

Figure 2.4. Nuclease activity of TALEN-AvrXa10 variants bearing N- and C-terminal extensions against target sites with the indicated spacers. TALEN scaffolds were indicated by the truncated N- and C-terminal positions as shown in **Figure 2.3**. Dot area indicates the relative β-GAL activity and represents the average of at least three independent experiments. "EBE" indicates that the cleavage target contains just one monomer binding site. The numeric values are shown in **Table 2.1**.

reduced activity against PTSs with a 24 bp spacer or a 32 bp spacer.

To further enhance the nuclease activity of the TALENs, we sought to add extra C-terminal extensions on the N3-C0 variant, which showed the highest DNA cleavage efficiency for all the tested substrates. The N3-C1 variant containing a 63 amino acid C-terminal extension cleaved longer DNA substrates (24 and 32 bp spacers) with >3-fold increased efficiency compared with N3-C0 variant, which allowed the TALEN to target PTSs in a moderately wide range (14-32 bp). We reasoned that a longer linker between

Table 2.1. Relative β-GAL activity of TALEN-AvrXa10 variants bearing N- and C-terminal extensions against target sites with the indicated spacers. The activity of I-CreI was normalized to 1. Values represent the average (± standard deviation) of at least three independent experiments. "EBE" indicates that the cleavage target contains just one monomer binding site.

	EBE	**6bp**	**8bp**	**10bp**	**12bp**	**14bp**	**15bp**	**16bp**	**24bp**	**32bp**
N0-C0	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
N1-C0	ND	ND	ND	1.7 ± 0.4	2.7 ± 0.7	2.3 ± 0.2	2.4 ± 0.4	2.6 ± 0.2	0.2 ± 0.0	ND
N2-C0	ND	ND	ND	2.1 ± 0.2	2.8 ± 0.1	2.5 ± 0.1	2.3 ± 0.6	3.2 ± 0.2	0.6 ± 0.1	0.2 ± 0.0
N3-C0	ND	ND	ND	2.7 ± 0.7	2.6 ± 0.7	2.2 ± 0.4	3.0 ± 0.4	3.1 ± 0.4	0.9 ± 0.2	0.3 ± 0.1
N4-C0	ND	ND	ND	2.4 ± 0.2	3.0 ± 0.1	1.9 ± 0.4	2.7 ± 0.3	2.2 ± 0.5	0.4 ± 0.2	0.2 ± 0.0
N3-C1	ND	ND	ND	ND	0.1 ± 0.0	1.6 ± 0.3	2.9 ± 0.7	3.1 ± 0.8	3.1 ± 0.5	3.1 ± 0.3
N3-C2	ND	ND	ND	ND	ND	0.5 ± 0.1	0.7 ± 0.5	2.1 ± 0.8	2.3 ± 0.5	1.9 ± 0.2
N3-C3	ND	ND	ND	ND	ND	ND	0.1 ± 0.0	0.8 ± 0.2	0.6 ± 0.1	0.3 ± 0.1
N3-C4	ND	ND	ND	ND	ND	ND	0.1 ± 0.0	1.5 ± 0.4	0.6 ± 0.0	0.4 ± 0.0
N3-C5	ND	ND	ND	ND	ND	0.1 ± 0.0	0.2 ± 0.0	0.5 ± 0.4	1.6 ± 0.1	1.5 ± 0.6

ND: no detectable activity

the TALE repeat units and the FokI cleavage domain may allow greater flexibility for the nuclease domain to cut close to or away from the ends of the EBE. However, additional extended C-terminal segments as shown in the N3-C2, N3-C3, N3-C4 and N3-C5 variants led to reduced nuclease activity. Compared to N3-C0, longer C-terminal extensions abolished the nuclease activity against PTSs with 10 bp and 12 bp spacers, indicating that longer protein linkers require longer spacers to accommodate the additional length of the peptide.

Taken together, TALENs harboring a short C-terminal extension (31 amino acids) require at least a 125 amino acid N-terminal fragment (as in the N1-C0 variant) to allow for efficient DNA cleavage against PTSs with 10-16 bp spacers, while the N3-C1 and N3-C2 variants with longer C-terminal extensions are capable of cleaving PTSs with longer spacers. Furthermore, none of the TALEN variants are able to cleave single EBE

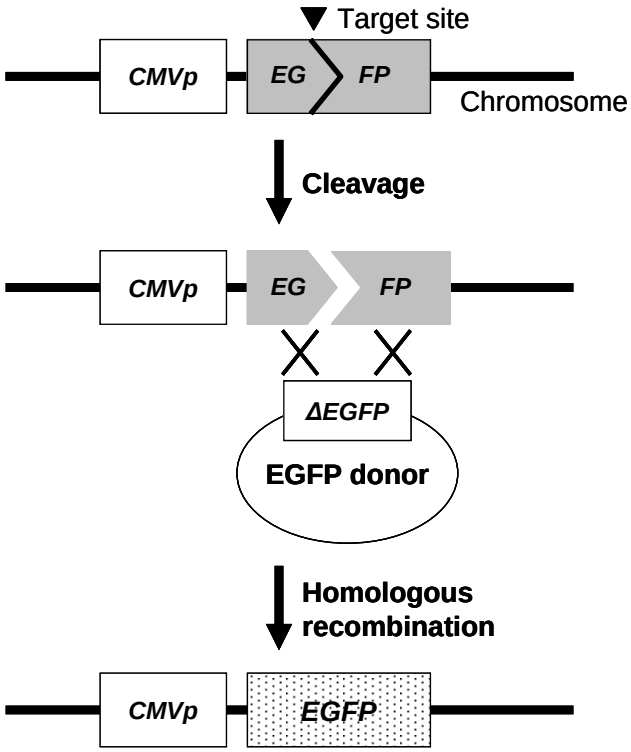

Figure 2.5. Schematic of the mammalian gene targeting system. Details are provided in the text. Recovered *EGFP* gene was shown in dotted box.

site or PTSs with 6 bp and 8 bp spacers, indicating an optimal range of spacer lengths (10-32 bp) between the two EBE sites is essential for efficient TALEN cleavage as a homodimer.

2.2.2. Custom-designed TALENs stimulated gene targeting in human cells

After demonstrating efficient DNA cleavage by engineered TALENs in yeast, we next sought to show that TALENs can produce a chromosomal DSB and modify the human genome. With a modified The enhanced green fluorescence protein (*EGFP*) gene conversion system (22), we were able to rapidly and quantitatively gauge the potential of TALEN-driven genome editing (**Figure 2.5**). The *EGFP* gene was divided into two

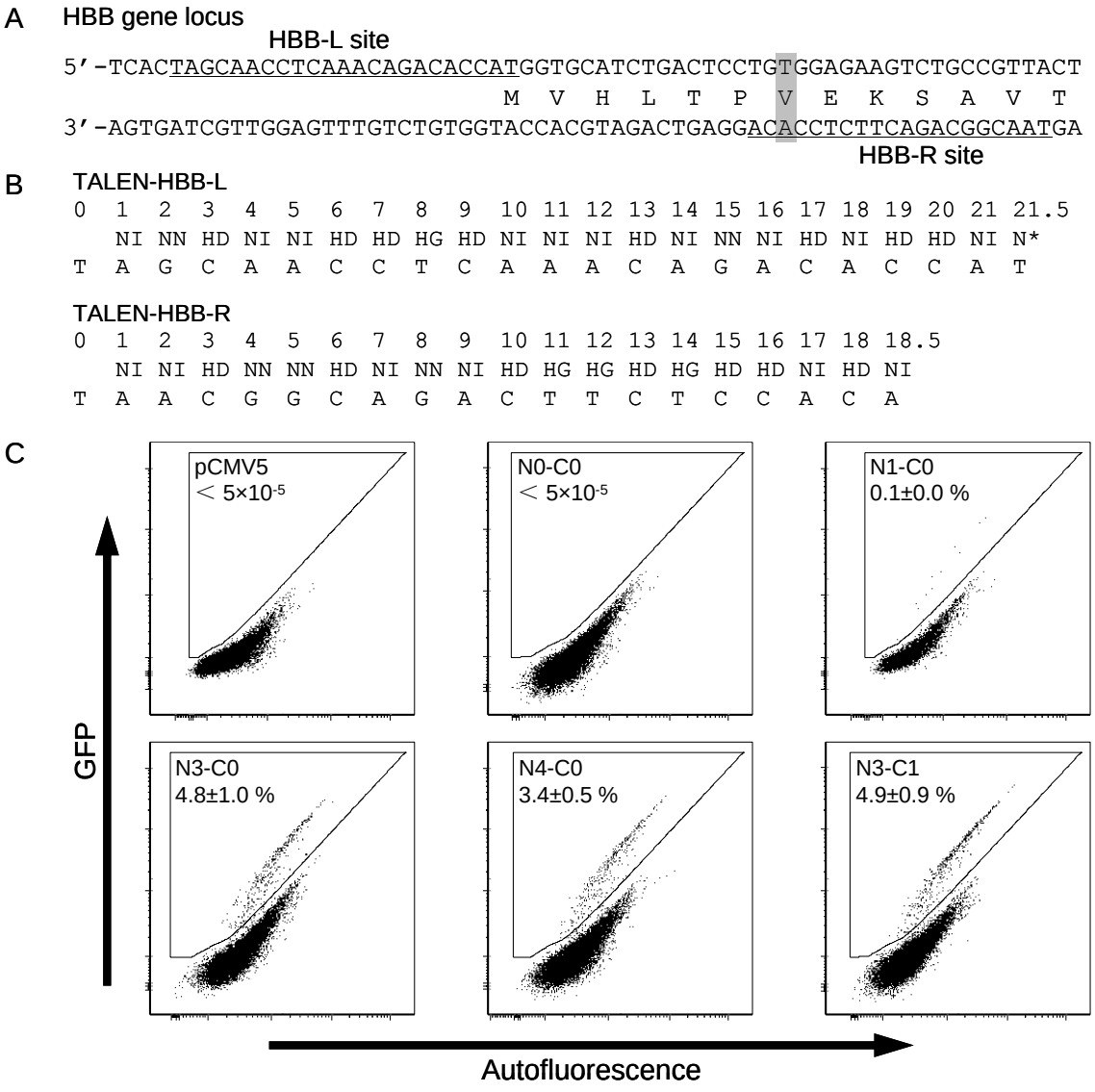

Figure 2.6. Custom-designed TALENs stimulated gene targeting in human cells. (**A**) Sequence of the *HBB* gene locus containing the single mutation (from A to T) causing SCD (shaded in grey) and the two user-defined TALEN binding sites (underlined). Amino acid sequence of β-globin is shown between the two DNA strands. (**B**) RVDs of the two custom-designed TALENs with their DNA target sequences shown below. An asterisk indicates a deletion at residue 13. (**C**) Representative flow cytometry plots of gene targeting. GFP-positive cells were quantitated in gate as depicted. Autofluorescence was measured using phycoerythrin-fluorescence channel. Empty vector pCMV5 served as negative control. The values represent the average (± standard deviation) of the gene targeting rates from at least three independent experiments.

truncated and non-functional fragments by insertion of the TALEN target site in the middle of the gene, which was stably integrated into the genome of Hela cells. To perform the assay, the expression plasmids carrying a pair of TALEN genes were co-transfected into the reporter cell line together with the donor plasmid bearing a promoter-less and non-functional *EGFP* gene, which served as a template for repairing chromosomal DSB. Site-specific cleavage of the *EGFP* reporter gene by the TALEN pair will promote homology-dependent gene conversion, generating GFP-positive cells that can be quantified by flow cytometry.

Based on the above results of TALEN architecture optimization and guidelines for TALEN site selection (23), we constructed a pair of custom-designed TALENs (referred to as TALEN-HBB-L and TALEN-HBB-R) as a heterodimer to target the disease-causing mutational *HBB* gene locus (**Figure 2.6A & B**). The two target sites were separated by 15 bp. Cells that were transfected with an empty vector (pCMV5) or a minimal TALEN (N0-C0) had low autofluorescence or background recombination frequencies ($< 5\times10^{-5}$) (**Figure 2.6C**). Introduction of custom-designed TALENs with the N3-C0 or N3-C1 configuration converted 4.8% or 4.9% of the reporter cells to GFP-positive respectively, stimulating the gene targeting rate by over 1000-fold. This result is on par with or better than the reported ZFNs tested using a similar reporter system, which converted 2-5% cells to GFP-positive (22,24). The N4-C0 configuration had less activity compared with N3-C0 and N3-C1, converting 3.4% of the reporter cells to GFP-positive. Surprisingly, the N1-C0 construct showed only 0.1% gene targeting rate, even though this configuration was efficient in cleaving palindromic sites in the yeast reporter system. Western blot analysis indicated that the TALEN-HBB-L construct with the N1-C0

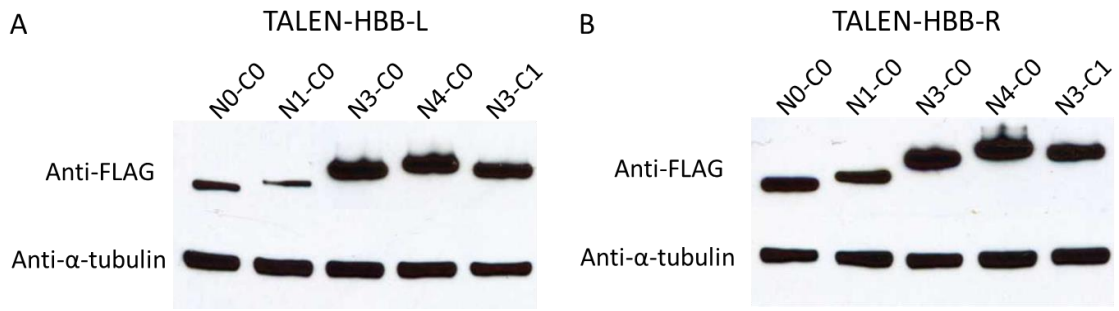

Figure 2.7. Western blot analysis for the expression of TALEN variants. (**A**) Expression level of TALEN-HBB-L with different configurations. (**B**) Expression level of TALEN-HBB-R with different configurations. All the TALEN variants contained a FLAG tag at the N-terminus. α-tubulin was used as an internal control.

configuration did not express well in human cells (**Figure 2.7A**), which may partially explain the lack of activity of this variant. On the other hand, all the tested TALEN-HBB-R variants including the N1-C0 variant exhibited good expression levels (**Figure 2.7B**). Based on the gene targeting system discussed above, we analyzed each TALEN's cleavage activity as a homodimer (**Figure 2.8** and **Figure 2.9**). TALEN-HBB-L with the N1-C0 configuration exhibited little activity (0.4%) against the HBB-L palindromic site partially due to a low expression level, which is consistent with the activity as a heterodimer. However, the low GFP signal (0.5%) from TALEN-HBB-R with the N1-C0 configuration indicated its low DNA cleavage activity, because it was expressed well in human cells. Similar to a heterodimer, each TALEN monomer bearing the N3-C0, N4-C0 or N3-C1 configuration allowed for efficient cleavage as a homodimer, exhibiting >2% gene targeting rate. Taken together, the TALEN pairs we created with at least 207 amino acid N-terminal segment were capable of recognizing and cleaving preselected sites on the *HBB* gene locus efficiently.

A HBB-L palindromic site

5′-TGTAGCAACCTCAAACAGACACCATACTGATCTAGGACTAATGGTGTCTGTTTGAGGTTGCTAGT

3′-ACATCGTTGGAGTTTGTCTGTGGTATGACTAGATCCTGATTACCACAGACAAACTCCAACGATCA

B

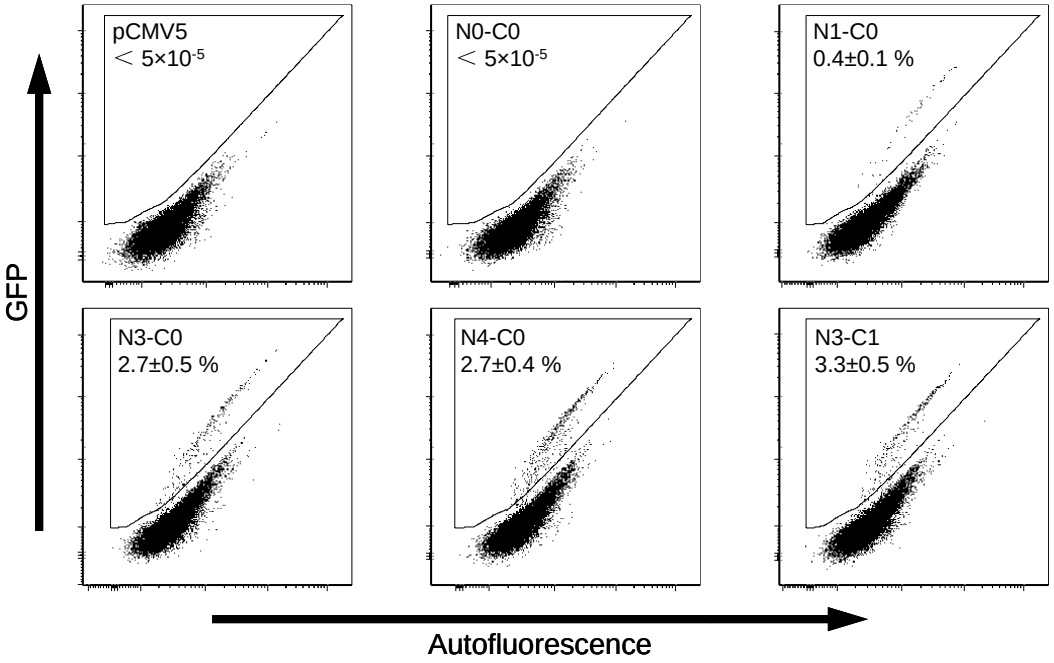

Figure 2.8. Activity of TALEN-HBB-L variants as a homodimer in a mammanlian gene targeting system. (A) Sequence of the HBB-L palindromic site with two HBB-L sites (underlined) in a tail-to-tail orientation separated by a 15 bp spacer. (B) Representative flow cytometry plots of gene targeting. GFP-positive cells were quantitated in gate as depicted. Autofluorescence was measured using phycoerythrin-fluorescence channel. Hela cells with an *EGFP* reporter gene integrated into the genome were co-transfected with the expression vectors of TALEN-HBB-L variants and an *EGFP* donor as the DNA repair template. Empty vector pCMV5 served as negative control. The values represent the average (± standard deviation) of gene targeting rates from at least three independent experiments.

A HBB-R palindromic site

5′-CCCTGTAACGGCAGACTTCTCCACAACTGATCTAGGACTATGTGGAGAAGTCTGCCGTTAGTGAA

3′-GGGACATTGCCGTCTGAAGAGGTGTTGACTAGATCCTGATACACCTCTTCAGACGGCAATCACTT

B

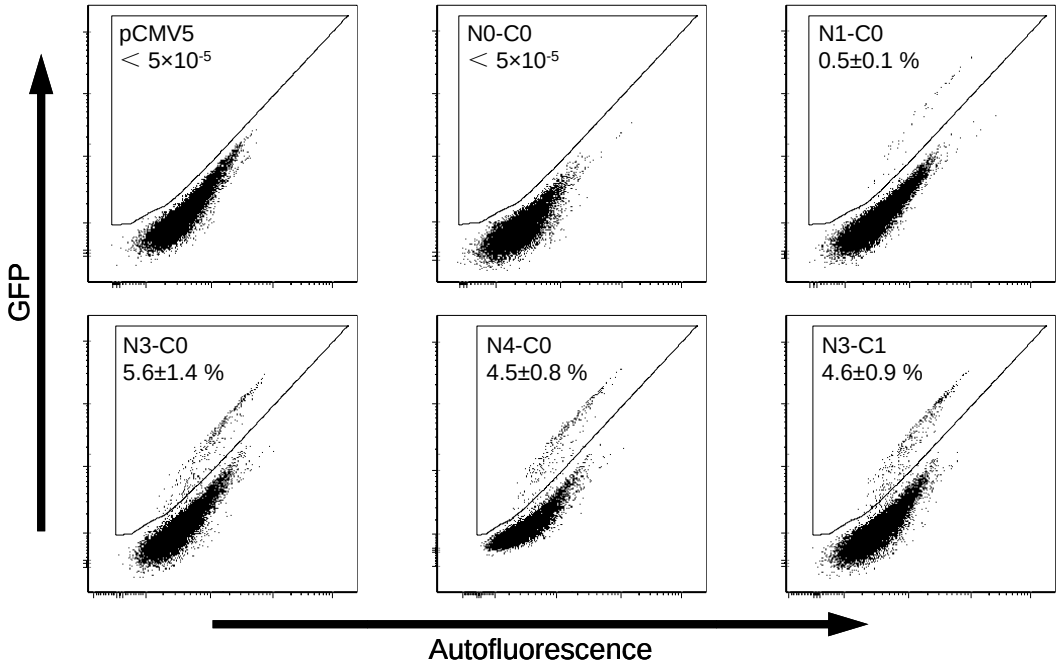

Figure 2.9. Activity of TALEN-HBB-R variants as a homodimer in a mammanlian gene targeting system. (A) Sequence of the HBB-R palindromic site with two HBB-R sites (underlined) in a tail-to-tail orientation separated by a 15 bp spacer. (B) Representative flow cytometry plots of gene targeting. GFP-positive cells were quantitated in gate as depicted. Autofluorescence was measured using phycoerythrin-fluorescence channel. Hela cells with an *EGFP* reporter gene integrated into the genome were co-transfected with the expression vectors of TALEN-HBB-R variants and an *EGFP* donor as the DNA repair template. Empty vector pCMV5 served as negative control. The values represent the average (± standard deviation) of gene targeting rates at least three independent experiments.

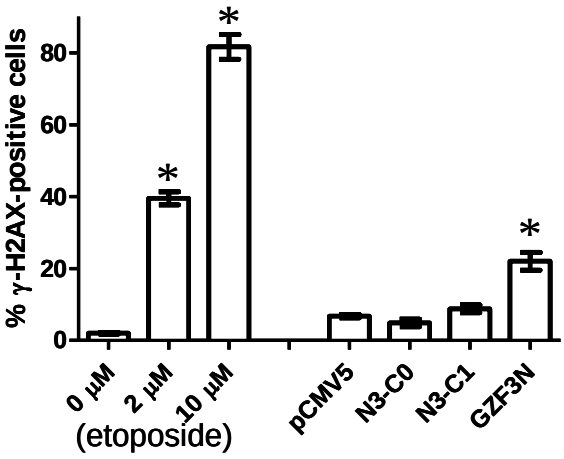

Figure 2.10. Nuclease-associated cytotoxicity of TALENs. The columns denote the percentage of γ-H2AX positive cells either transfected with artificial DNA nucleases or treated using etoposide with indicated concentration. Error bars indicate standard deviation from the average of at least three independent experiments. *, statistically significant increase of toxicity as compared to empty vector pCMV5 ($P < 0.001$).

2.2.3. Custom-designed TALENs caused no detectable cytotoxicity

The DNA binding specificity is a major determinant of cytotoxicity for ZFNs (25). Since the novel TALEN pair we designed recognizes a 43 bp long DNA recognition sequence, we postulated that it might not cause significant cytotoxicity due to its high specificity, even though there is some degeneracy in RVD recognition (i.e. NN can recognize G as well as A) (26). To assess the nuclease-associated toxicity, we monitored the intracellular level of γ-H2AX, which is the H2AX histone phosphorylated on serine 139 as a reaction on a chromosomal DSB (27). HEK293 cells displayed an average γ-H2AX level similar to the background (pCMV5 control) after being transfected with the TALEN pairs under the same conditions used to induce recombination in the gene targeting assay (**Figure 2.10**). In contrast, when we transfected the cells with a reported toxic ZFN (28) or treated

the cells with etoposide (an anti-cancer agent that causes DNA DSBs (29)) for 2 h, significantly higher levels of H2AX phosphorylation were detected. Therefore, the engineered TALENs caused little, if any, cytotoxicity in human cells, demonstrating their high specificity.

2.2.4. Targeting "unnatural" TALE sites

Naturally occurring TALE recognition sites are uniformly preceded by a 5'-T (17); therefore, all of the previously reported TALENs were designed to target DNA sequences bearing a T at position 0. To test whether a preceding T is an essential requirement for efficient cleavage by TALENs, we selected five TALEN-AvrXa10 variants to measure their nuclease activities towards "unnatural" TALE sites with a preceding A, C or G using the yeast reporter system. The spacer length between each monomer binding site was fixed to be 16 bp, allowing for efficient cleavage for all the tested TALEN variants (**Figure 2.4**). As shown in **Figure 2.11**, the N2-C0 and N3-C0 constructs harboring short C-terminal extensions exhibited moderate nuclease activity against unnatural PTSs (0.5-1 fold efficiency compared with I-CreI), which is still significantly lower than that towards the natural PTS with a T at position 0. The N4-C0 construct with a full length N-terminal segment displayed slightly reduced DNA cleavage efficiency towards all the tested substrates. The N3-C1 construct with additional C-terminal extension compared with the N3-C0 construct showed much higher cleavage efficiency against unnatural substrates. The N3-C1 construct cleaved PTSs bearing a preceding A, C or G with an efficiency similar to that for natural PTS with a preceding T. The extended 63 amino acid peptide linker between the DNA binding domain and the FokI cleavage domain in the N3-C1

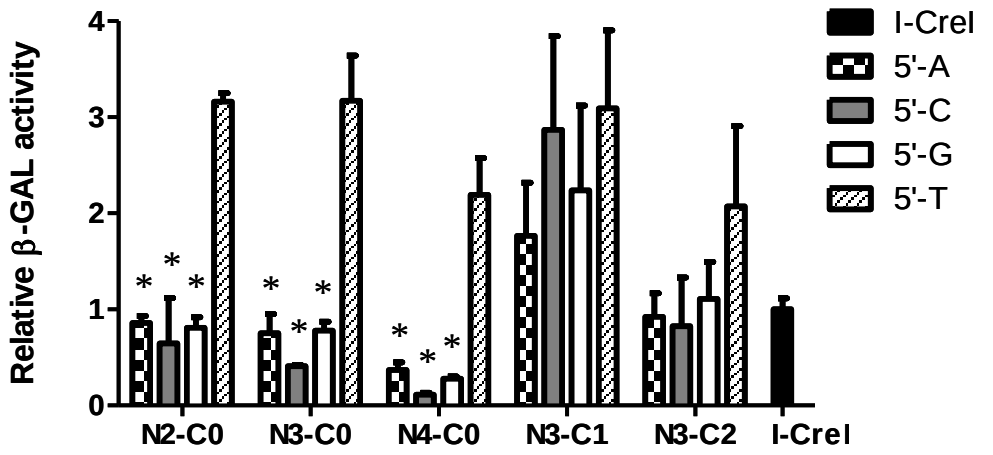

Figure 2.11. Cleavage efficiency of TALEN-AvrXa10 variants targeting unnatural TALE substrates. Two EBE sites preceded with A, C, G or T separated by a 16 bp spacer were used as DNA substrates. I-CreI served as an internal control and the nuclease activity was normalized to 1. The columns denote the relative β-GAL activity and represent the average of at least three independent experiments. Error bars indicate standard deviation. * indicates significantly decreased activity ($P<0.001$) compared to the 5'-T preceded substrate.

construct makes the TALEN less selective in terms of the 5' preceding nucleotide. Longer C-terminal extension in the N3-C2 construct reduced the nuclease activity, but still exhibited less selectivity for a 5'-T. The decreased selectivity for the preceding nucleotide exhibited by the N3-C1 scaffold provides more flexibility in choosing endogenous target loci for genome editing when designing TALENs.

2.3. Discussion

In this study, we constructed ten TALEN variants and determined the optimal range of spacer lengths between the two EBE sites for each variant to allow for efficient DNA cleavage. The original TALENs were constructed by fusing the FokI catalytic domain

with full length TALEs (30-32). Since the naturally occurring TALEs are transcription activators from a plant bacterial pathogen, their N-terminal segment contains signal peptides responsible for protein translocation and C-terminal segment harbors nuclei localization signals and transcription activator domain (33). These redundant fragments can be detrimental to the TALENs nuclease activity. Therefore, TALEN scaffold needs to be optimized to achieve efficient cleavage. Previously, Miller *et al.* (26) truncated the first 152 N-terminal residues from TALE and produced a set of TALENs with different C-terminal truncations to maximize nuclease activity. Their variants with 28 amino acid or 63 amino acid C-terminal extensions exhibited highest DNA cleavage efficiency, which are similar to our N3-C0 and N3-C1 constructs, respectively. We found that N3-C1 construct bearing 63 residues displayed efficient cleavage against DNA substrates with long spacers (16-32 bp), while Miller *et al.* did not observe this activity, which may be due to the shorter N-terminal segment in their TALEN variants. Even though they reported undetectable nuclease activity of TALEN with 93 amino acid C-terminal fragment, we observed efficient cleavage using N3-C4 and N3-C5 variants against DNA substrates with longer spacers, which is in agreement with the original study on TALENs bearing full length TALE C-terminus. More recently, Mussolino *et al.* (34) reported the most active configuration was a TALEN bearing 153 amino acid N-terminal segment and 17 or 47 amino acid C-terminal segment, cleaving target sites with 12-15 bp spacer. However, their *in vivo* reporter system relied on the disappearance of GFP signal, in which the binding of TALENs to the target sites without cleavage may generate false positive signals (e.g., some of their TALEN variants exhibited positive result targeting single EBE site). Because the different TALEN activities in the above mentioned reports

were determined using different reporter systems, it is difficult to compare their results directly with ours. However, our systematic study clearly identifies two optimized TALEN scaffolds, with N3-C0 variant preferring PTSs with shorter (10-16 bp) spacers and N3-C1 variant preferring PTSs with longer (14-32 bp) spacers. This indicates that our optimized TALEN architectures have a moderately wide optimal range of spacer length between each TALEN monomer binding site, which allows for more flexibility with the choice of recognition sites in gene targeting study.

The single mutation in *HBB* gene causes sickle cell disease, against which there is no adequate long-term treatment. Gene-insertion based gene therapy has been reported to be effective in mouse model by introduction of an antisickling *HBB* gene copy (3,4). However, immunogenicity and insertional mutagenesis caused by integrating viral vectors are still under safety concerns, which limit the wide application of this technology. Therefore, it is tempting to correct the mutational *HBB* gene *in situ*. In this study, we constructed a pair of TALENs that were able to recognize a user-defined sequence within *HBB* gene locus and stimulated gene targeting rate by over 1000-fold without detectable cytotoxicity. The TALENs with N3-C0 and N3-C1 scaffolds displayed highest cleavage efficiency, which is consistent with our result from yeast reporter system. To be noticed, N1-C0 variant exhibited high nuclease activity against PTS with 15bp spacer in the yeast reporter system but showed little activity in the human gene targeting system, even though Western blot analysis indicated its good expression level. The inconsistency may be due to the difference between episomal cleavage in our yeast reporter system and chromosomal cleavage in the more complex genome in human cells. Taken together, our optimized TALENs with N3-C0 or N3-C1 scaffold exhibited robust

nuclease activity and low cytotoxicity in human cells, which can serve as a powerful tool for gene repair in gene therapy.

A 5'-T preceding the TALE EBE sites was believed to be essential for TALE function (17), which, based on our study, turned out not to be the case for TALE nucleases. All the tested TALEN variants exhibited good nuclease activity against unnatural DNA substrate (0.5-1 fold efficiency compared with I-CreI); even though TALENs variants with short C-terminal extension still preferred to cleave natural TALEN recognition site. Interestingly, N3-C1 TALEN variant was capable of cleaving unnatural DNA substrates preceded by 5'-A, C or G with similar efficiency compared with that for natural TALE site. However, the longer C-terminal extension in N3-C1 variant is at the 3'-end of the EBE, which is physically distant from the 5'-end of target site. We reasoned that the presence of a leucine zipper-like heptad repeats closely linked to the TALE central repeat units (35) may increase the DNA binding affinity of the central repeat domain and thus make the 5'-T interaction less of a requirement. More detailed structural studies will be needed to elucidate the protein-DNA interface of TALENs and their corresponding DNA substrates in order to solve this uncertainty. Our result indicates that the requirement for a 5'-T can be mitigated using certain scaffolds, thus allowing greater flexibility in choosing desired target site in genome editing endeavors.

2.4. Materials and methods

2.4.1. Materials

Phusion DNA polymerase, T4 DNA ligase, T4 polynucleotide kinase, antarctic phosphatase, and restriction endonucleases were purchased from New England Biolabs (Beverly, MA). QIAprep Spin Plasmid Miniprep Kit, QIAquick Gel Extraction Kit, and QIAquick PCR Purification Kit were obtained from Qiagen (Valencia, CA). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). All the other reagents unless specified were obtained from Sigma-Aldrich (St Louis, MO).

2.4.2. Yeast reporter system

The yeast reporter system was constructed based on a single strand annealing assay with slight modifications (20). The *lacZ* gene was divided into two truncated and non-functional copies by insertion of a palindromic target site consisting of two AvrXa10 EBEs in a tail-to-tail orientation with various spacers. The two separated *lacZ* fragments sharing a 100 bp homologous region were PCR amplified separately, digested with *Avr*II and *Xba*I respectively and ligated overnight at 16 °C. The ligation product was subsequently digested with a combination of *Avr*II and *Xba*I and the 3.3 kb non-cut fragment was gel purified. The *lacZ* reporter gene was then cloned into the pRS414 plasmid in between the *GAL1* promoter and the *ADH1* terminator, whose expression was induced by adding 2% galactose in the culture medium.

2.4.3. Construction of TALEN expression vectors

The DNA sequence encoding the nonspecific DNA cleavage domain of FokI restriction endonuclease was PCR-amplified from pST1374 (plasmid 13426; Addgene, Cambridge, MA). The PCR product was inserted into pET28a(+) vector (EMD, Madison, WI) through *Nde*I and *Sal*I recognition sites to yield plasmid pET28a-G4S-FokI. The DNA sequence encoding the central repeat DNA binding domain of TALE AvrXa10 was isolated from 0.7% agarose gel after digestion of the plasmid pAS202 (kindly provided by Dr. Frank F. White from Kansas State University, Manhattan, KS) using restriction enzymes *Stu*I and *Aat*II. The DNA fragment after gel extraction was cloned into the *Stu*I and *Aat*II sites of plasmid pET28a-G4S-FokI to make pET28a-TALEN-Xa10. The DNA sequences encoding the N-terminal extensions of the AvrXa10 protein with varying length were PCR-amplified from pAS202. The PCR products were digested with *Nde*I and *Stu*I and cloned into the large fragment of the pET28a-TALEN-Xa10 vector digested with the same enzymes. The DNA sequences encoding the C-terminal extensions of the AvrXa10 protein with varying length were PCR-amplified using the same template and cloned into the *Aat*II and *Afl*II sites of the plasmid pET28a-TALEN-Xa10. The gene encoding TALEN-Xa10 with different configurations were subcloned into pRS415 plasmid through the *Nde*I and *Sal*I sites in between the GAL10 promoter and the ADH2 terminator. The expression of TALENs in *Saccharomyces cerevisiae* was induced by adding 2% galactose in the cultural medium.

The DNA sequences encoding the custom-designed TALE central repeat domains targeting the specific sequence within the HBB gene locus were synthesized by DNA2.0, Inc (Menlo Park, CA). The synthesized DNA sequences were codon optimized to

increase the protein expression level and reduce the repetitiveness among amino acid repeats in the DNA binding domain to facilitate DNA cloning. The genes of custom-designed TALENs with different N- and C-terminal extensions were first cloned into the pET28a(+) vector using the same strategy as described above. Next, the TALEN genes along with a FLAG tag sequence and a SV40 nucleus localization signal added to the N-terminus were subcloned into the pCMV5 mammalian expression vector (36) through the *Kpn*I and *Sal*I sites. The expression of these constructs is under the control of the cytomegalovirus (CMV) promoter.

To test the *in vivo* cleavage activity of TALENs in the yeast reporter system, both the pRS414 reporter plasmid and pRS415 expression plasmid were transformed into *Saccharomyces cerevisiae* HZ848 (*MATα*, *ade*2-1, *ade*3Δ22, *Δ*ura3, *his*3-11,15, *trp*1-1, *leu*2-3,112 and *can*1-100) using the LiAc/SS carrier DNA/PEG method (37). After 3-4 days growth at 30 °C on plates containing synthetic complete medium lacking leucine and tryptophan (SC-Leu-Trp) with 2% glucose, three transformants were picked to grow in SC-Leu-Trp liquid medium with 2% galactose. After cultivation for 2 days at 30 °C, yeast cells were harvested by centrifugation and lysed by Y-PER yeast protein extraction reagent (Pierce, Rockford, IL). The β-galactosidase (β-GAL) enzyme activity was determined by the β-galactosidase enzyme system from Promega (Madison, WI) according to the manufacturer's instruction. For each sample, the β-GAL activity was measured as a function of time and normalized by total protein concentration determined using the BCA protein assay (Pierce, Rockford, IL). Homing endonuclease I-CreI (38) serves as an internal control and its activity was normalized to one.

2.4.4. Human gene targeting system

EGFP gene was divided into two fragments by insertion of a preselected sequence from the *HBB* gene locus and an in-frame stop codon. The non-functional *EGFP* reporter gene was cloned into the pLNCX2 retroviral vector (BD Clontech, Palo Alto, CA) under the cytomegalovirus (CMV) promoter and stably integrated into the genome via retroviral transduction according to the manufacturer's protocol. Transfected Hela cells were selected in 500 μg/ml G418 for two weeks. A donor plasmid was constructed by insertion of a promoter-less *EGFP* gene with the first 37 nucleotides missing into the pNEB193 plasmid (New England Biolabs, Beverly, MA) through the *Hind*III and *Sal*I sites. The donor plasmid provides a homologous DNA segment as a template for repairing the DSB in the *EGFP* reporter. Construction of TALEN expression vectors is described in the Supplementary Data.

Reporter Hela cells were maintained in modified Eagle's medium (MEM) supplemented with 10% Fetal Bovine Serum (FBS; Hyclone, Logan, UT). Cells were seeded in 12-well plates at a density of 1×10^5 per well. After 24 hr, reporter cells were co-transfected with 333 ng of each TALEN expression plasmid and 333 ng of donor plasmid using FuGene HD transfection reagent (Promega, Madison, WI) under conditions specified by the manufacturer. The transfection efficiency was determined to be between 45% and 50%. Cells were trypsinized from their culturing plates 48h after transfection and resuspended in 200 μL phosphate buffered saline (PBS) supplemented with 10 mM EDTA for flow cytometry analysis. 20,000 cells were analyzed by BD LSRII flow cytometer (BD Biosciences, San Jose, CA) to determine the percentage of

GFP-positive cells. The rate of gene targeting was determined by normalizing the percentage of GFP-positive cells to the transfection efficiency.

2.4.5. Immunoblotting

Hela cells in 12-well plates were harvested 24 h after transfection with TALEN variants. Cells were collected by centrifugation, washed with PBS and resuspended in 40 μL RIPA lysis buffer (Santa Cruz Biotechnology, Santa Cruz, CA) with SDS-PAGE loading dye. Proteins were resolved by 4-20% Mini-Protean TGX Precast Gel (Bio-Rad Laboratories, Hercules, CA), transferred onto a nitrocellulose membrane, blocked for 1 hr with Tris-buffered saline/0.05% Tween 20 containing 5% nonfat milk, followed by incubation with anti-FLAG tag (1:500) and anti-α-tubulin (1:5000) antibodies at 4 °C overnight. After incubation with anti-mouse horseradish peroxidase-conjugated secondary antibody (1:25000; GeneScript, Piscataway, NJ) for 1 hr, bands were visualized using SuperSignal West Pico Chemiluminescent Substrate (Pierce, Rockford, IL).

2.4.6. H2AX phosphorylation assay

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS. Cells were seeded in 12-well plates (2×10^5 /well) and transfected after 24 hr with 500 ng of each TALEN expression plasmid, 1 μg empty pCMV5 vector as a negative control or 1 μg reported toxic ZFN construct GZF3N (kindly provided by Dr. Toni Cathomen of Hannover Medical School, Hannover, Germany) as a positive control, respectively. After 48 h post transfection, cells were harvested, fixed, permeabilized, and stained using the H2AX phosphorylation assay kit

(Millipore, Watford, UK) according to the manufacturer's protocol. Alternatively, cells were treated with dimethyl sulfoxide (DMSO), 2 μM and 10 μM etoposide in DMSO for 2 hr before staining. Cells were then scanned in a flow cytometer to quantitate the number of cells staining positive for phosphorylated histone H2AX.

2.5. References

1. Ashley-Koch, A., Yang, Q. and Olney, R.S. (2000) Sickle hemoglobin (HbS) allele and sickle cell disease: a HuGE review. *Am J Epidemiol*, **151**, 839-845.
2. Frenette, P.S. and Atweh, G.F. (2007) Sickle cell disease: old discoveries, new concepts, and future promise. *J Clin Invest*, **117**, 850-858.
3. Levasseur, D.N., Ryan, T.M., Pawlik, K.M. and Townes, T.M. (2003) Correction of a mouse model of sickle cell disease: lentiviral/antisickling beta-globin gene transduction of unmobilized, purified hematopoietic stem cells. *Blood*, **102**, 4312-4319.
4. Pawliuk, R., Westerman, K.A., Fabry, M.E., Payen, E., Tighe, R., Bouhassira, E.E., Acharya, S.A., Ellis, J., London, I.M., Eaves, C.J. *et al.* (2001) Correction of sickle cell disease in transgenic mouse models by gene therapy. *Science*, **294**, 2368-2371.
5. Marshall, E. (1999) Gene therapy death prompts review of adenovirus vector. *Science*, **286**, 2244-2245.
6. Check, E. (2002) A tragic setback. *Nature*, **420**, 116-118.
7. Chang, J.C., Ye, L. and Kan, Y.W. (2006) Correction of the sickle cell mutation in embryonic stem cells. *Proc Natl Acad Sci U S A*, **103**, 1036-1040.

8. Thomas, K.R. and Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, **51**, 503-512.

9. Bollag, R.J., Waldman, A.S. and Liskay, R.M. (1989) Homologous recombination in mammalian cells. *Annu Rev Genet*, **23**, 199-225.

10. Cathomen, T. and Joung, J.K. (2008) Zinc-finger nucleases: the next generation emerges. *Mol Ther*, **16**, 1200-1207.

11. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, **11**, 636-646.

12. Carroll, D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther*, **15**, 1463-1468.

13. Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods*, **5**, 374-375.

14. Kim, J.S., Lee, H.J. and Carroll, D. (2010) Genome editing with modularly assembled zinc-finger nucleases. *Nat Methods*, **7**, 91; author reply 91-92.

15. Radecke, S., Radecke, F., Cathomen, T. and Schwarz, K. (2010) Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol Ther*, **18**, 743-753.

16. Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, **14**, 49-55.

17. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.

18. Hopkins, C.M., White, F.F., Choi, S.H., Guo, A. and Leach, J.E. (1992) Identification of a family of avirulence genes from *Xanthomonas-oryzae* pv. *oryzae*. *Mol Plant-Microbe Interact*, **5**, 451-459.

19. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10570-10575.

20. Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*, **31**, 2952-2962.

21. Gurlebeck, D., Szurek, B. and Bonas, U. (2005) Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. *Plant J*, **42**, 175-187.

22. Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.

23. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res*, **39**, e82.

24. Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646-651.

25. Cornu, T.I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J.K. and Cathomen, T. (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther*, **16**, 352-358.

26. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, **29**, 143-148.

27. Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S. and Bonner, W.M. (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem*, **273**, 5858-5868.

28. Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol*, **25**, 786-793.

29. Kaufmann, S.H. (1989) Induction of endonucleolytic DNA cleavage in human acute myelogenous leukemia cells by etoposide, camptothecin, and other cytotoxic anticancer drugs: a cautionary note. *Cancer Res*, **49**, 5870-5878.

30. Bikard, D., Julie-Galau, S., Cambray, G. and Mazel, D. (2010) The synthetic integron: an in vivo genetic shuffling device. *Nucleic acids research*, **38**, e153.

31. Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) Modularly assembled designer TAL effector nucleases

for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*, Epub ahead of print 31 March 2011; doi: 2010.1093/nar/gkr2188.

32. Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A*, **108**, 2623-2628.

33. White, F.F., Potnis, N., Jones, J.B. and Koebnik, R. (2009) The type III effectors of Xanthomonas. *Mol Plant Pathol*, **10**, 749-766.

34. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, **39**, 9283-9293.

35. Yang, Y. and Gabriel, D.W. (1995) Xanthomonas avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. *Mol Plant Microbe Interact*, **8**, 627-631.

36. Andersson, S., Davis, D.L., Dahlback, H., Jornvall, H. and Russell, D.W. (1989) Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J Biol Chem*, **264**, 8222-8229.

37. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, **2**, 31-34.

38. Chevalier, B.S., Monnat, R.J., Jr. and Stoddard, B.L. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol*, **8**, 312-316.

CHAPTER 3. SEAMLESS GENE CORRECTION OF SICKLE CELL DISEASE MUTATION IN HUMAN INDUCED PLURIPOTENT STEM CELLS USING TALENS

3.1. Introduction

Human induced pluripotent stem cells (hiPSCs) are genetically reprogrammed from adult somatic cells. Similar to human embryonic stem cells (hESCs), hiPSCs can propagate indefinitely in culture and differentiate to all somatic tissues (1). Disease-specific hiPSCs derived from patients have invaluable therapeutic applications such as studying disease mechanisms, screening effective drugs, and perhaps most importantly, treating diseases via cell replacement therapies (2,3). For the treatment of genetic disorders which are caused by genomic abnormalities, patient-derived hiPSCs can be corrected in culture and thus provide a renewable cell source for autologous transplantations, in which modified hiPSCs (or their differentiated progenies) can be transplanted back into the patient to restore normal cell functions (4).

To correct genetic defects in hiPSCs, a gene-insertion based strategy has been developed in which viral vectors are used to carry a normal gene for the substitution of the defective endogenous counterpart (5). However, immunogenicity and insertional mutagenesis caused by integrating viral vectors are major safety concerns. Alternatively, the mutated gene can be corrected *in situ* by homologous recombination (HR), which not only circumvents the safety concerns but also ensures the corrected gene expressed in the natural genetic context and thus in the appropriate temporal and tissue-specific manner.

However, the low HR frequency ($<10^{-6}$) in human stem cells prevents its wide application (6).

To address this limitation, custom-designed DNA endonucleases such as zinc finger nucleases (ZFNs) and transcription activator-like (TAL) effector nucleases (TALENs) have been constructed to introduce a site-specific DNA double-strand break (DSB) on the chromosome and enhance HR frequency by >1000 fold (see **Chapter 1**). ZFNs have been successfully applied to induce HR-based gene correction in patient-derived hiPSCs (7-10). Yet the low specificity of the ZFNs can cause cleavage at off-target sites, resulting in undesired genomic instability (11,12). Moreover, the limited modularity of zinc finger domains makes it difficult for ZFNs to target many disease-causing genes. In contrast, TALENs represent a new class of artificial nucleases with high specificity and modularity (13,14). The central repeat domain of a TALEN consists of repeating units of 33-35 amino acids. Each repeat recognizes a single nucleotide and the specificity is conferred by the highly variable di-residues at positions 12 and 13 (*e.g.* Asn-Ile recognizes A, His-Asp recognizes C, Asn-Gly recognizes T, and Asn-Asn recognizes G and A) (15,16). The simple DNA recognition code and the modular nature make TALENs an ideal platform for use in targeted gene correction.

To explore the potential therapeutic application of TALENs in hiPSC-based cell and gene therapies, we chose sickle cell disease (SCD) as a model system (see **Chapter 2**). SCD is the most common autosomal recessive disorder that affects more than 300 000 individuals worldwide each year (17). A homozygous mutation (from A to T) in the sixth codon of the human β-globin (*HBB*) gene converts a glutamate to a valine, which generates abnormal β-globin proteins and results in malfunctional red blood cells (18).

The absence of an adequate long-term treatment makes hiPSCs-based cell replacement therapy highly attractive. In this study, we successfully applied TALENs to correct the SCD mutation in patient-derived hiPSCs. In combination with *piggyBac* transposon, TALEN-mediated gene correction left no residual ectopic sequences at the site of correction. Treated hiPSCs maintained the genomic integrity and the pluripotency. This study provides an important proof of principle that TALENs can be used in hiPSC-based cell and gene therapies. Similar strategies can be applied to correct other genetic disorders, which represents a significant advance in basic biomedical research and human clinical applications.

3.2. Results

3.2.1. Construction of a donor plasmid for *HBB^S* gene correction

We constructed a pair of custom-designed TALENs to recognize the sickle *HBB* (*HBB^S*) gene at the mutation site (**Figure 2.6**) (19). In order to correct the sickle mutation (E6V) of the *HBB^S* gene by TALEN-induced HR, we constructed a donor plasmid to provide an overall ~2.5 kb homology sequence carrying a wild-type *HBB* gene (*HBB^A*) with GAG for codon six to substitute GTG (**Figure 3.1**). The donor plasmid also contains a *piggyBac* transposon, which is a mobile element that transposes efficiently in human stem cells (9). The *piggyBac* transposon carries a bifunctional fusion protein (puroΔtk) constitutively expressed under PGK promoter, in which puro (puromycin *N*-acetyltransferase) is used for positive selection and Δtk (a truncated version of herpes simplex virus type 1 thymidine kinase) is used for negative selection (20). The P_{PGK}-puroΔtk cassette is flanked by two *piggyBa*c terminal inverted repeats (PB5 and PB3)

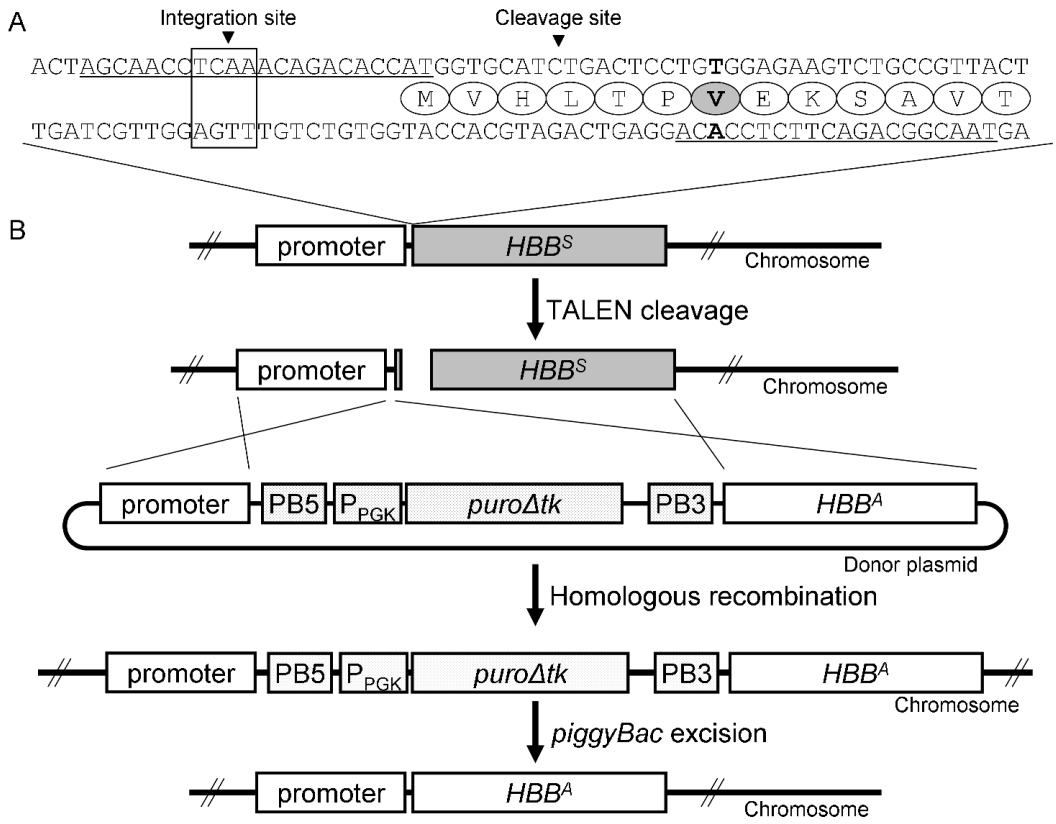

Figure 3.1. Schematic overview depicting the strategy for the seamless correction of the *HBB*^{S} gene using TALENs and *piggyBac* transposon. (**A**) Sequence of the genomic *HBB*^{S} gene locus. Amino acids of the beginning part of human β^S-globin protein are shown in ovals. The A to T substitution (bold letter) causes the E6V mutation (grey oval) in human β^S-globin. The two TALEN recognition sites (underlined) are separated by a 15 bp spacer. The TALEN cleavage site and the integration site for *piggyBac* transposon (in frame) are indicated. (**B**) Schematic of the donor plasmid design and gene correction strategy. The donor plasmid comprises two homology arms and a *piggyBac* transposon (dotted).

into the TTAA site to enable its subsequent excision. Expression of *piggyBac* transposase will excise the transposon seamlessly, without leaving any residual sequences.

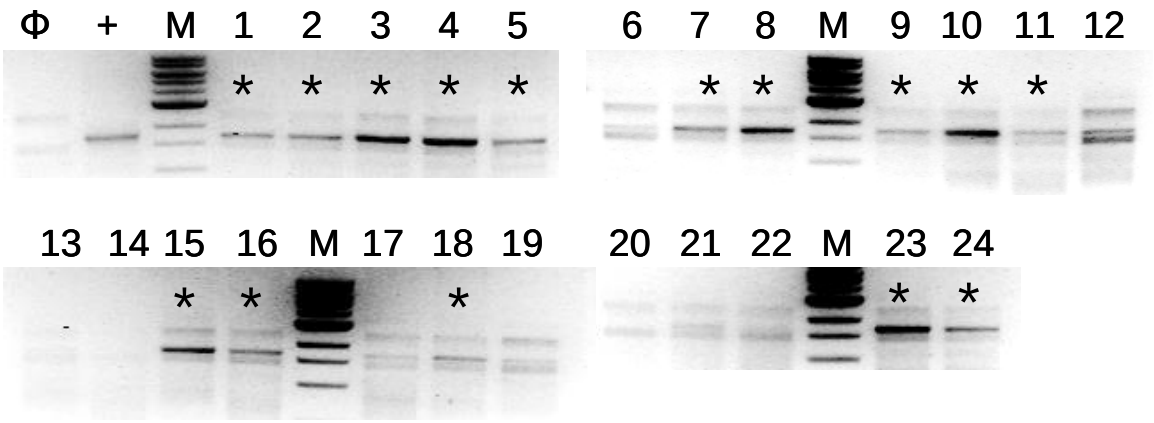

Figure 3.2. Representative results of nested-PCR screening of puromycin-resistant colonies after TALEN-induced HR. Φ, genomic PCR product from untreated hiPSCs as a negative control; +, internal round of PCR product from the donor plasmid as a positive control; M, 1 kb DNA Ladder (New England Biolabs, Beverly, MA); *, genomic PCR product from correctly targeted hiPSC colonies.

3.2.2. Seamless correction of the *HBB*^{S} gene in the SCD patient-derived hiPSCs

To correct the *HBB*^{S} gene, we transfected SCD patient-derived hiPSCs carrying a homozygous E6V mutation with the TALEN-expression plasmids together with the donor plasmid. TALEN cleavage induced HR, which enabled the replacement of the chromosomal *HBB*^{S} gene with the episomal *HBB*^{A} gene and simultaneously integrated a *piggyBac* transposon carrying the drug-selectable cassette (**Figure 3.1B**). Puromycin-resistant hiPSC colonies were picked and expanded for nested-PCR screening. We observed that >50% puromycin-resistant colonies were correctly targeted (**Figure 3.2**). To remove the drug-selectable cassette from the modified hiPSCs, the *piggyBac* transposase gene was transiently transfected to the cells. After negative selection under fialuridine, hiPSC colonies were picked and expanded. Genotyping was performed by genomic PCR followed by restriction digestion with *Bsu*36I, which cleaves the corrected

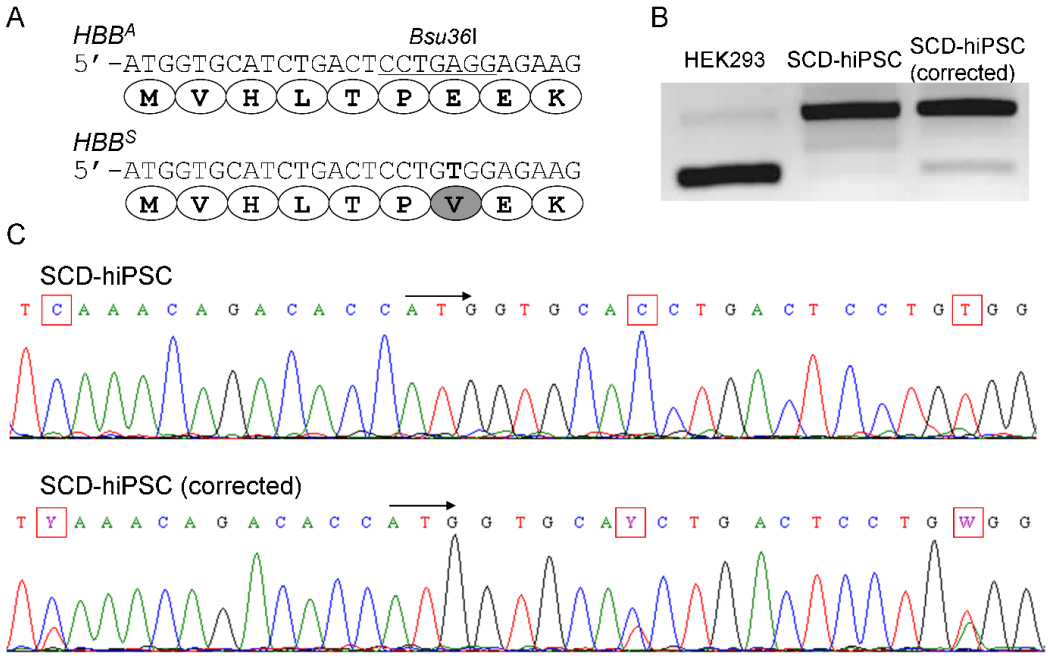

Figure 3.3. Seamless monoallelic gene correction of the *HBB*^{S} gene in SCD patient-derived hiPSCs. (**A**) Sequences of sickle (*HBB*^{S}) and wild-type (*HBB*^{A}) human β-globin gene. Amino acids from the beginning part of human β-globin protein are shown in ovals. The A to T substitution (bold letter) causes the E6V mutation (grey oval) of *HBB*^{S} gene, which abolishes a *Bsu*36I restriction site in *HBB*^{A} gene. (**B**) *Bsu*36I analysis confirming the correction of the *HBB*^{S} gene. HEK293 cell line has both *HBB*^{A} alleles, which are sensitive to *Bsu*36I digestion. SCD patient-derived hiPSCs (SCD-hiPSC) have both *HBB*^{S} alleles, which are resistant to *Bsu*36I digestion. Corrected SCD-hiPSCs have one *HBB*^{S} allele and one *HBB*^{A} allele, as indicated by the double bands. (**C**) Sequence analysis showing the *HBB* gene locus of the untreated patient-derived hiPSCs and the gene-corrected hiPSCs. The ATG start codon is indicated by a black arrow. The nucleotide substitutions are highlighted by red frames. Y, double peaks of C and T; W, double peaks of A and T.

(*HBB*^{A}) but not the sickle (*HBB*^{S}) gene (**Figure 3.3A & B**). Two out of 48 hiPSC colonies were sensitive to *Bsu*36I digestion, indicating the correction of the mutation on one allele (**Figure 3.4**). Sequencing of the genomic PCR products of the two corrected

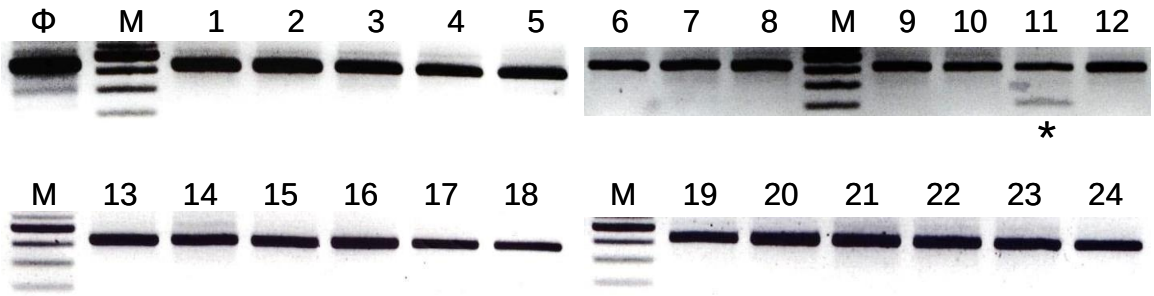

Figure 3.4. Representative results of *Bsu*36I digestion screening of fialuridine-resistant hiPSC colonies after excision of the *piggyBac* transposon. Φ, analysis of the untreated hiPSCs as a negative control; M, 100 bp DNA Ladder (New England Biolabs); *, correct hiPSC colony with one HBB^S allele corrected and the *piggyBac* transposon removed.

hiPSC colonies confirmed that transposon excision did not leave any residual "scar", demonstrating that one HBB^S gene in the SCD patient-derived hiPSCs was corrected seamlessly (**Figure 3.3C**). Two introduced silent mutations were observed as initially designed, confirming that the replacement was introduced by TALEN-mediated gene correction, and not by spontaneous reversion.

3.2.3. Characterization of the gene-corrected hiPSCs

The gene-corrected hiPSCs maintained pluripotency, as indicated by the uniform expression of pluripotency-specific marker proteins including OCT4, SOX2, SSEA4, TRA-1-60 and TRA-1-81 (**Figure 3.5**). To test pluripotency *in vivo*, the corrected hiPSCs were transplanted into severe combined immunodeficiency (SCID) mice and tumor formation was observed in 9-10 weeks. Histological examination showed that the tumor comprised cell types originating from all three developmental germ layers (**Figure 3.6**), including gland (endoderm), duct (endoderm), cartilage (mesoderm), blood vessel

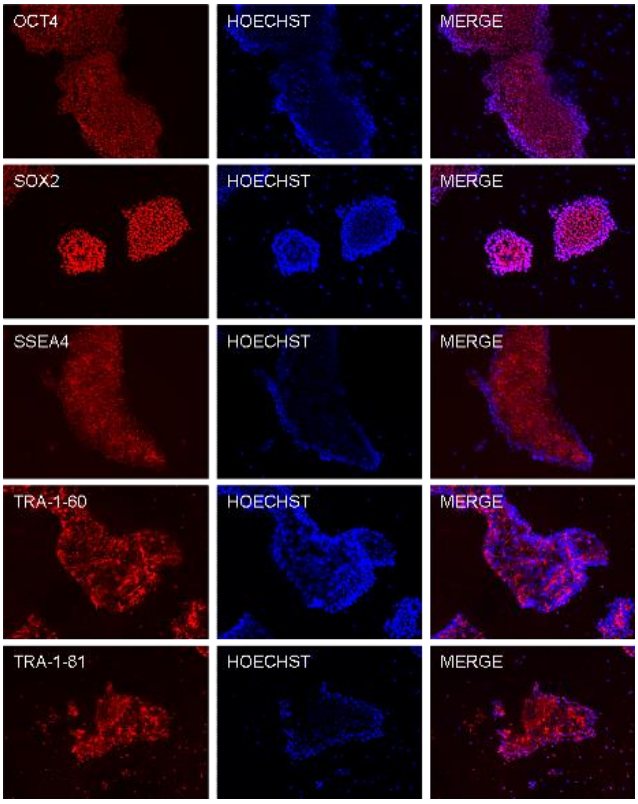

Figure 3.5. Immunostaining analysis. The corrected hiPSCs display characteristic pluripotency markers including OCT4, SOX2, SSEA4, TRA-1-60 and TRA-1-81. Nuclei were stained with HOECHST.

(mesoderm), white adipose tissue (mesoderm) and neural rosette (ectoderm). Those evidences indicate that TALEN-mediated genome editing did not alter the pluripotent state of the corrected hiPSCs.

To monitor possible off-target events introduced by TALEN cleavage, we analyzed six genomic regions harboring the potential off-target cleavage sites (predicted by TAL Effector Nucleotide Targeter 2.0 (21)), including a highly similar sequence within the *HBD* gene that shares only four mismatches compared with the designed TALEN binding site (**Table 3.1**). Off-target activity was measured by SURVEYOR nuclease assay (Material and methods), which monitors the small insertions and deletions generated by

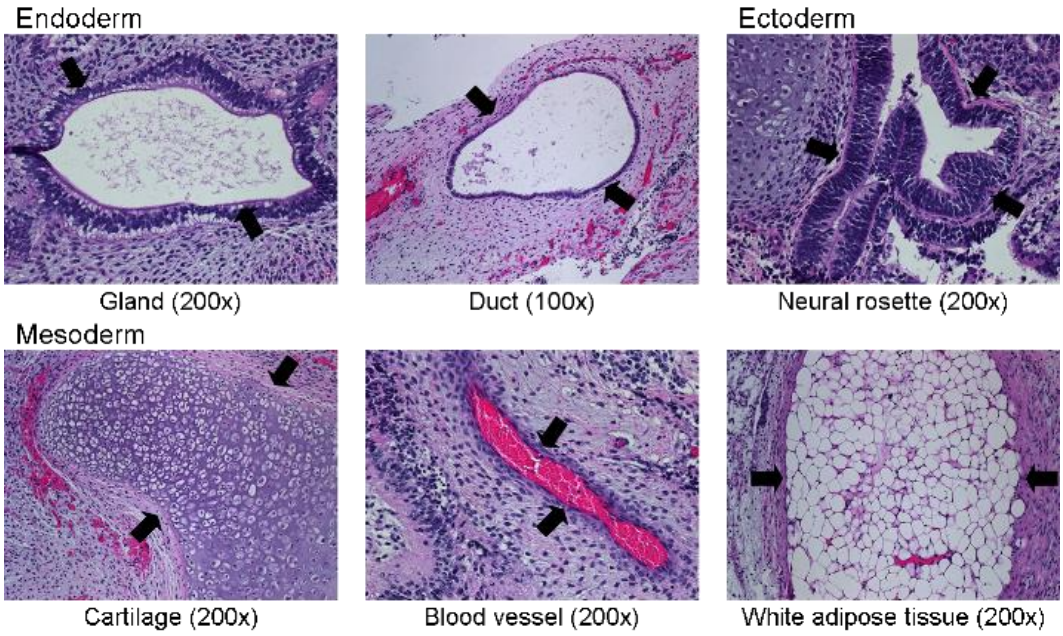

Figure 3.6. Teratoma formation and analysis of the corrected hiPSCs. Hematoxylin and eosin staining indicates *in vivo* differentiation of endodermal (gland and duct), ectodermal (neural rosette) and mesodermal (cartilage, blood vessel and white adipose tissue) structures.

imprecise DNA repair from non-homologous end joining (NHEJ) in response to TALEN cleavage. In contrast to *HBB* gene locus, within which exists three mismatches between the corrected (HBB^{A}) and uncorrected (HBB^{S}) genes, none of the other six analyzed sites were off-targeted (**Figure 3.7A**). In addition, karyotype analysis showed that the TALEN-mediated gene correction and *piggyBac* transposon excision processes maintained genomic integrity of the corrected hiPSCs without causing any unexpected chromosomal translocations and alterations (**Figure 3.7B**). These evidences indicate high specificity and safety of using TALENs and *piggyBac* transposon for targeted genome editing in patient-derived hiPSCs.

Table 3.1. Potential off-target sites of the custom-designed TALENs for the recognition of *HBB*^{S} gene. Bases differing from the left TALEN recognition site (underlined) and the right TALEN recognition site (bold) are highlighted. The possible off-target events within *ZNF204P* and *CDH18* genes are generated by the formation of homodimer of the right TALEN.

Featured Gene	TALEN recognition sites		Mismatch (bp)		Spacer length (bp)
	Left	Right	Left	Right	
***HBB*^{S}**	AGCAACCTCAAACAGACACCAT	AACGGCAGACTTCTCCACA	0	0	15
HBD	AGCAACCTCAAACAGACACCAT	**G**AC**A**GCAG**T**CTTCTCC**T**CA	0	4	15
HLF	AGGGAACCCAAACAATCACCAT	AAC**A**GCA**AA**G**C**TC**C**CCA**G**A	6	6	14
PTPRG	AGCAACCTGAAACACTCCCACT	**T**AC**A**G**AA**T**A**CTTCTC**A**AC**T**	6	6	16
NSUN5	AAAAACCCCAAACCCACACCCA	A**C**C**A**G**T**AGAC**C**TCT**G**CAC**C**	7	6	27
ZNF204P	AAC**T**G**A**AG**G**CTT**T**TCCACA	AAC**AT**CAGA**TAA**CTC**A**ACA	4	6	29
CDH18	**C**A**TAT**CA**C**ACTT**T**TCCACA	**C**A**A**G**C**CA**C**ACT**A**C**A**CCACA	6	6	30

3.3. Discussion

Obtainable from adult somatic cells and capable of growing indefinitely and differentiating into all cell types, hiPSCs open up various novel avenues for basic biomedical research and human therapeutics. The ultimate promise of hiPSCs, although at a very early stage of development, is to treat diseases by cell replacement therapy. Patient-derived hiPSCs can provide a renewable autologous cell source for the treatment of degenerative diseases such as central nervous system injuries, myocardial infarction, and diabetes (22). Moreover, if the disease-causing mutations can be repaired *in vitro*, corrected hiPSCs can serve as regenerative medicines for the treatment of genetic disorders such as SCD and β-thalassemia. Toward this end, a proof-of-principle has been established in a SCD mouse model, including derivation of disease-specific mouse induced pluripotent stem cells (miPSCs) from adult somatic cells, repair of genetic mutations by HR-mediated gene targeting, differentiation of corrected miPSCs into therapeutic relevant cells, and subsequent autologous transplantation (4). However,

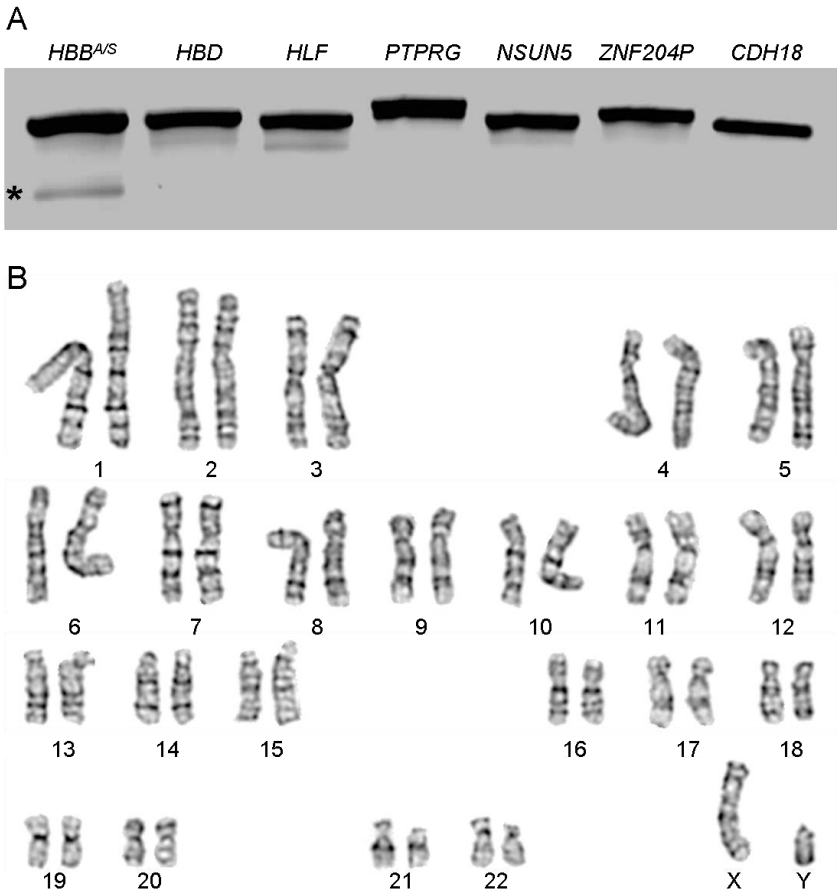

Figure 3.7. Analysis of off-target effects. (**A**) SURVEOR nuclease analysis of the potential off-target sites. The *HBB*$^{A/S}$ site has three mismatches between corrected (*HBB*^{A}) and uncorrected (*HBB*^{S}) genes, which is sensitive to SURVEOR nuclease cleavage. All of the six potential off-target sites (sequences are shown in **Table 3.1**) are resistant to SURVEOR nuclease cleavage. *, product of SURVEOR nuclease cleavage. (**B**) Karyotyping analysis of the corrected hiPSCs.

similar approaches are not applicable to human gene therapy because of the low HR rate in human cells, making the gene correction step a major obstacle.

To address this limitation, rare-cutting DNA endonucleases such as ZFNs and TALENs have been constructed to introduce a chromosomal DSB close to the mutation site. Subsequent DSB repair by HR can increase the gene targeting rate significantly (23-

25). ZFNs have been applied to correct disease-causing mutations in hiPSCs derived from patients with SCD (7,10), Parkinson's disease (8), and α_1-antitrypsin deficiency (9). However, the lack of specificity of zinc finger domains raises safety concerns about potential off-target cleavage, which might lead to undesired mutations and chromosomal aberrations (11,12). In addition, zinc finger domains have limited modularity due to the context-dependent DNA-binding effects. Recently, TALENs have rapidly emerged as an alternative to ZFNs and have been widely applied for genome editing in many different organisms and cell types (19,26-38). The modular nature of the TALE central repeat domains enables researchers to tailor DNA recognition specificity with ease and target essentially any desired DNA sequence (15,16). Moreover, TALENs exhibit minimal off-target effects as determined by exome sequencing and whole-genome sequencing (37), indicating their potential utility for human therapeutics. In this study, we sought to explore the therapeutic potential of TALENs in hiPSCs-based cell and gene therapies. We successfully applied TALENs to correct the SCD mutation in patient-derived hiPSCs. The corrected hiPSCs maintain their pluripotent state. TALEN-mediated gene correction does not introduce detectable off-target events or chromosomal alterations.

Although the chromosomal DSBs introduced by custom-designed DNA endonucleases can significantly stimulate HR, the overall gene targeting rate is still low, which makes it necessary to introduce a drug-selectable cassette into the chromosome for the isolation of targeted colonies. To remove the unwanted drug-selection cassette after positive selection, the Cre/*loxP* recombination system has been frequently used, which leaves a 34 bp *loxP* sequence at the site of excision (7,10,35,39,40). However, the small ectopic sequence has the potential to silence the gene expression of the targeted allele (10)

or interfere with transcription regulatory elements of adjacent genes (41). To address this limitation, we inserted the drug-selectable cassette into a *piggyBac* transposon, which enables the removal of the selection cassette efficiently in hiPSCs without leaving any remnant sequences (9). After TALEN-mediated gene correction and drug selection, transient expression of *piggyBac* transposase excised the drug-selectable cassette without leaving any residual "scar" in the corrected gene. The combination of TALEN and *piggyBac* technology enabled us to correct the SCD mutation in patient-derived hiPSCs in a seamless manner.

Because robust protocols for differentiating hiPSCs into mature, enucleated red blood cells *in vitro* have not been established, functional data demonstrating the recovery of the gene-corrected red blood cells cannot be generated in this study. In addition, a reproducible *in vivo* model is needed to test the efficacy and safety of using the gene-corrected hiPSCs as a renewable autologous cell source for the treatment of SCD patients (10,42,43). Although challenges and barriers remain, our study demonstrates an important first step of using TALEN technology and *piggBac* transposon to correct the disease-causing mutation *in situ* in a seamless manner, which represents a significant advance toward hiPSC-based cell and gene therapies.

3.4. Materials and methods

3.4.1. Cell culture

SCD patient-derived hiPSCs with a homozygous E6V mutation were kindly provided by Dr. Marius Wernig from Stanford University (Stanford, CA). Cultures were maintained under feeder-free culture conditions on 6-well plates coated with Matrigel Basement

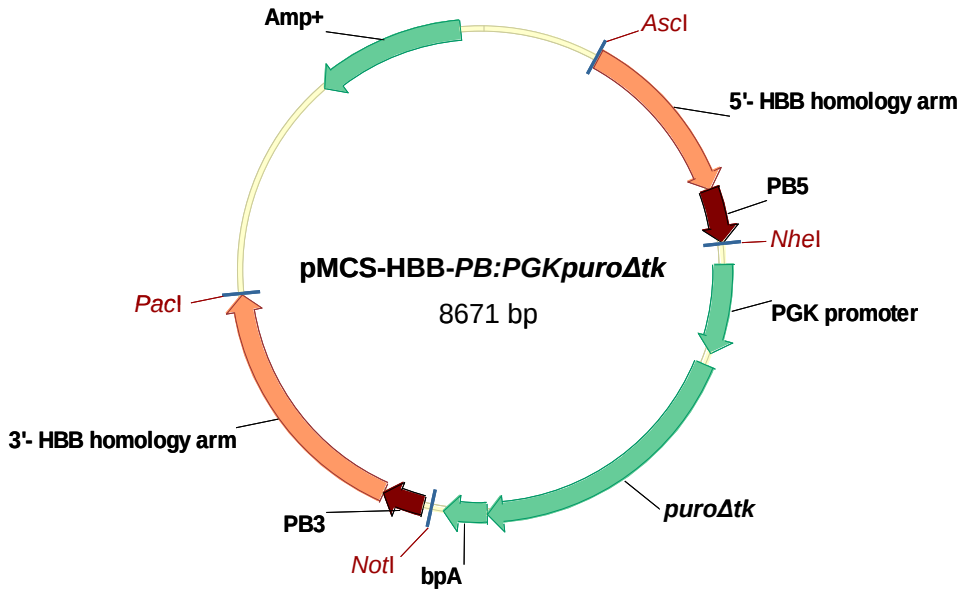

Figure 3.8. Plasmid map of the donor plasmid.

Membrane Matrix (BD Biosciences, Bedford, MA) in mTeSR1 medium (StemCell Technologies, Vancouver, Canada). Subculture was performed every 5 to 7 days by washing hiPSCs with DMEM/F12 medium, incubating with 1 mg/mL dispase (StemCell Technologies) for 7 min at 37 °C, washing with DMEM/F12 medium, detaching with a cell scraper, breaking down into small clumps and plating onto a new Matrigel-coated plate.

3.4.2. Plasmid construction

Human expression plasmids for the TALEN pair targeting the *HBB*^{S} gene have been described previously (19). The donor plasmid pMCS-HBB-*PB:PGKpuroΔtk* (**Figure 3.8**) was constructed by replacing the 5'- and 3'- homology arms of pMCS-AAT-*PB:PGKpuroΔtk* (9) (kindly provided by Dr. Allan Bradley from Wellcome Trust Sanger Institute, Cambridge, UK) with homology sequences of the *HBB*^{A} gene. The 5'- and 3'-

HBB homology arms were amplified by PCR from the genomic DNA of HEK293 cells. The primers AscI-HBB-5'-for (5'-act gac *ggc gcg cc*t att ctt agt gga cta gag g-3', *Asc*I site shown in italic) and PB5-HBB-5'-rev (5'-gca tgc gtc aat ttt acg cag act atc ttt cta ggg tta agg ttg cta gtg aac aca gtt-3') amplified the 1.0 kbp 5'- *HBB* homology arm. The primers PB3-HBB-3' (5'-ttt aac gta cgt cac aat atg att atc ttt cta ggg tta aac aga cac cat ggt gca tct-3') and PacI-HBB-3'-rev (5'-act gac *tta att aa*t gga cag caa gaa agc gag c-3', *Pac*I site shown in italic) amplified the 1.5 kbp 3'- *HBB* homology arm. The two terminal inverted repeats (PB5 and PB3) of *piggyBac* transposon were amplified by PCR from pMCS-AAT-*PB:PGKpuroΔtk*. PB5 was amplified by primers HBB-5'-PB5-for (5'-ctt aca ttt gct tct gac aca act gtg ttc act agc aac ctt aac cct aga aag ata gtc-3') and NheI-PB5-rev (5'-act tgt *gct agc* att cta gtt gat atc tat aac-3', *Nhe*I site shown in italic). PB3 was amplified by primers NotI-PB3-for (5'-gct aga *gcg gcc gc*a ctc gag ata tct aga c-3', *Not*I site shown in italic) and HBB-3'-PB3-rev (5'-cag act tct cct cag gag tca gat gca cca tgg tgt ctg ttt aac cct aga aag ata atc-3'). 5'-*HBB* homology arm was combined with PB5 by overlap extension-PCR (OE-PCR) (44) and inserted into *Asc*I and *Nhe*I sites of pMCS-AAT-*PB:PGKpuroΔtk* to form pMCS-Intermediate. 3'-*HBB* homology arm was combined with PB3 by OE-PCR and inserted into *Not*I and *Pac*I sites of pMCS-Intermediate to form donor plasmid pMCS-HBB-*PB:PGKpuroΔtk*.

3.4.3. TALEN-mediated genome editing of hiPSCs

On day 0, SCD-hiPSCs were pre-treated with 2 μM of a Rho Kinase (ROCK) inhibitor (thiazovivin, BioVision, Milpitas, CA) for at least 1 hr prior to transfection. Cells were dissociated into single-cell suspension by washing with DMEM/F12, incubating with

accutase (StemCell Technologies) for 7 min at 37 °C, mixing with mTeSR1 medium followed by vigorous pipetting. 3×10^6 cells were transfected with 2 μg of each TALEN-encoding plasmid and 2 μg of donor plasmid by Amaxa nucleofection system using Human Stem Cell Nucleofector Kit 1 (Lonza, Cologne, Germany) according to the manufacturer's instruction. The treated cells were plated onto one Matrigel-coated 100 mm dish in mTeSR1 medium supplemented with 2 μM of thiazovivin for the first 24 hrs. On day 3, puromycin (Sigma, St. Louis, MO) selection at 0.5 μg/mL was started. Individual colonies were picked and expanded in one or two Matrigel-coated 24-well plates on day 13-17. On day 20-24, near-confluent cells in each well were harvested using accutase. One quarter of the cells was seeded onto a new Matrigel-coated 24-well plate and the rest of the cells were used for nested PCR analysis.

3.4.4. Nested PCR of targeted integration

Genomic DNA was extracted from hiPSCs using QuickExtract DNA Extraction Solution 1.0 (EPICENTRE Biotechnologies, Madison, WI) according to the manufacturer's instructions. The external round of PCR was carried out using primers Purotest-for-PB3 (5'-ggc tcg aga tcc act agt tc-3') and Purotest-rev-HBB3-long (5'-gat gct caa ggc cct tca ta-3'). The external round of PCR products were used as template for the internal round of PCR using primers Nested-for-HBB3 (5'-acg tgg atg aag ttg gtg gt-3') and HBB-R (5'-taa ggg tgg gaa aat aga cc-3'). Nested PCR products were then resolved on a 1.0% agarose gel.

3.4.5. Transposon excision in targeted hiPSCs

On day 0, correctly targeted hiPSCs were pre-treated with 2 μM of thiazovivin for at least 1 hr prior to transfection. Cells were dissociated into single-cell suspension as described earlier. 2×10^6 cells were transfected with 4 μg of the hyperactive *piggyBac* transposase expression vector (pCMV-hyPBase (45), kindly provided by Dr. Allan Bradley) by Amaxa nucleofection system using Human Stem Cell Nucleofector Kit 1 according to the manufacturer's instructions. The treated cells were plated onto one Matrigel-coated 100 mm dish in mTeSR1 medium supplemented with 2 μM of thiazovivin for the first 24 hrs. On day 3, cells were dissociated into single-cell suspension and seeded to a new Matrigel-coated 100 mm dish at low density (~1000 cells/mL) in mTeSR1 medium containing 2 μM of thiazovivin and 0.25 μM of fialuridine (Santa Cruz Biotechnology, Santa Cruz, CA). On day 6, medium was changed to mTeSR1 medium supplemented with 2.5 μM of fialuridine, which was used for daily medium change until picking colonies. Individual colonies were picked and expanded in one or two Matrigel-coated 24-well plates on day 16-20. On day 23-27, near-confluent cells in each well were harvested using accutase. One quarter of the cells was seeded onto a new Matrigel-coated 24-well plate and the rest of the cells were used for *Bsu*36I analysis and sequencing.

3.4.6. *Bsu*36I analysis and sequencing of genomic *HBB* locus

Genomic DNA from corrected hiPSCs was amplified with primers HBB-F (5'-agg tac ggc tgt cat cac tt-3') and HBB-R (5'-taa ggg tgg gaa aat aga cc-3') under standard PCR conditions, which resulted in a 402 bp product. PCR products were purified by QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA) and subjected to restriction digestion with

Table 3.2. Sequences of primers that are used to amplify the potential off-target sites for the SURVEYOR nuclease assay.

Gene	Primer sequences	
HBB	Forward	5'-agg tac ggc tgt cat cac tt-3'
	Reverse	5'-taa ggg tgg gaa aat aga cc-3'
HBD	Forward	5'-agg gaa tag tgg aat gaa gg-3'
	Reverse	5'-aga cca cca gta atc tga gg-3'
HLF	Forward	5'-gtc ata tgt tga aag agg gc-3'
	Reverse	5'-aga cca ttc caa ata tag cag g-3'
PTPRG	Forward	5'-ctg tat gta tcc aac atc ata ctg-3'
	Reverse	5'-ctc tgc agt tct gtc aag ga-3'
NSUN5	Forward	5'-gca ctg gaa tgg cta aag ag-3'
	Reverse	5'-tag tag cgg aga cag taa gg-3'
ZNF204P	Forward	5'-gtc tgt cct ctg cct aaa gg-3'
	Reverse	5'-aac tca cac tgg aga gaa gc-3'
CDH18	Forward	5'-acc atc aga tct cct gag ac-3'
	Reverse	5'-taa gtc taa gtg cgc tca tg-3'

*Bsu*36I (New England Biolabs, Beverly, MA) to detect the E6V mutation and subsequently separated on a 2% agarose gel. The E6V mutation of β-globin disrupts an existing *Bsu*36I restriction site, so that the digestion of the corrected allele results in two ~200 bp fragments whereas the mutant allele results in one 402 bp fragment. The correction event was identified by sequencing of the 402 bp PCR amplicons using HBB-F and HBB-R as sequencing primers.

3.4.7. Off-target cleavage analysis

Potential off-target cleavage sites were identified by TAL Effector Nucleotide Targeter 2.0 using high score cutoff (21). DNA fragments (~400 bp) containing the potential off-target sites were PCR-amplified from the genomic DNA extracted from hiPSCs using the primers described in **Table 3.2**. Gel-purified PCR products were reannealed slowly and

analyzed by SURVEYOR Mutation Detection Kits (Transgenomic, Omaha, NE) according to the manufacturer's instructions. Off-target cleavage of TALENs can be repaired by NHEJ, which typically generates deletions and insertions. Therefore, genomic PCR products from TALEN-modified region are heterogeneous. After reannealing, DNA heteroduplexes can be cleaved by SURVEYOR nucleases, which indicates off-target cleavage of the TALENs. Because the potential off-target sites are designed to be in the center of the genomic PCR products (~400 bp), cleavage by SURVEYOR nucleases can generate two ~200 bp fragments. On the contrary, genomic PCR products from TALEN-unmodified regions are homogeneous, which are resistant to SURVEYOR nuclease.

3.4.8. Immunostaining

Expression of the five pluripotent markers including OCT4, SOX2, SSEA4, TRA-1-60 and TRA-1-80 were analyzed by Human ES/iPS Cell Characterization Kit (Applied StemCell, Menlo Park, CA) according to the manufacturer's instruction. HOECHST stain was used to stain nuclei.

3.4.9. Karyotyping

Karyotyping of the gene-corrected hiPSCs was carried out by Applied StemCell Inc. (Menlo Park, CA).

3.4.10. Teratoma formation and analysis

The gene-corrected hiPSCs were dissociated with dispase and triturated as small clumps

and injected into kidney capsule and testis sites (1.5-3×10^6 cells per site) of three 6-week old male mice (Fox Chase SCIDR-beige from Charles River Laboratories, Wilmington, MA). Cells from a hiPSC line were injected into the same sites of one mouse as a positive control. Tumor formation was observed in the three mice and harvested at day 74 post injection. The tumor samples at the six sites were fixed in 10% formalin overnight, embedded in paraffin, cut into 5-μm serial sections, and analyzed by hematoxylin and eosin staining.

3.5. References

1. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861-872.
2. Saha, K. and Jaenisch, R. (2009) Technical challenges in using human induced pluripotent stem cells to model disease. *Cell Stem Cell*, **5**, 584-595.
3. Yamanaka, S. (2007) Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell*, **1**, 39-49.
4. Hanna, J., Wernig, M., Markoulaki, S., Sun, C.W., Meissner, A., Cassady, J.P., Beard, C., Brambrink, T., Wu, L.C., Townes, T.M. *et al.* (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*, **318**, 1920-1923.
5. Raya, A., Rodriguez-Piza, I., Guenechea, G., Vassena, R., Navarro, S., Barrero, M.J., Consiglio, A., Castella, M., Rio, P., Sleep, E. *et al.* (2009) Disease-corrected

haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature*, **460**, 53-59.

6. Zwaka, T.P. and Thomson, J.A. (2003) Homologous recombination in human embryonic stem cells. *Nat Biotechnol*, **21**, 319-321.
7. Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F. *et al.* (2011) In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*, **29**, 1717-1726.
8. Soldner, F., Laganiere, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, **146**, 318-331.
9. Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordonez, A., Hannan, N.R., Rouhani, F.J. *et al.* (2011) Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, **478**, 391-394.
10. Zou, J., Mali, P., Huang, X., Dowey, S.N. and Cheng, L. (2011) Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood*, **118**, 4599-4608.
11. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods*, **8**, 765-770.

12. Radecke, S., Radecke, F., Cathomen, T. and Schwarz, K. (2010) Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol Ther*, **18**, 743-753.

13. Sun, N. and Zhao, H. (2013) Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol Bioeng*, doi: 10.1002/bit.24890.

14. Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, **14**, 49-55.

15. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.

16. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.

17. Weatherall, D.J. and Clegg, J.B. (2001) Inherited haemoglobin disorders: an increasing global health problem. *Bull World Health Organ*, **79**, 704-712.

18. Orkin, S.H. and Higgs, D.R. (2010) Medicine. Sickle cell disease at 100 years. *Science*, **329**, 291-292.

19. Sun, N., Liang, J., Abil, Z. and Zhao, H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst*, **8**, 1255-1263.

20. Chen, Y.T. and Bradley, A. (2000) A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells. *Genesis*, **28**, 31-35.

21. Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K. and Bogdanove, A.J. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res*, **40**, W117-122.

22. Goldring, C.E., Duffy, P.A., Benvenisty, N., Andrews, P.W., Ben-David, U., Eakins, R., French, N., Hanley, N.A., Kelly, L., Kitteringham, N.R. *et al.* (2011) Assessing the safety of stem cell therapeutics. *Cell Stem Cell*, **8**, 618-628.

23. Jasin, M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet*, **12**, 224-228.

24. Smih, F., Rouet, P., Romanienko, P.J. and Jasin, M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res*, **23**, 5012-5019.

25. Sun, N., Abil, Z. and Zhao, H. (2012) Recent advances in targeted genome engineering in mammalian systems. *Biotechnol J*, **7**, 1074-1087.

26. Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S. and Zhang, B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol*, **29**, 699-700.

27. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, **29**, 143-148.

28. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome

editing activity in combination with low toxicity. *Nucleic Acids Res*, **39**, 9283-9293.

29. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, **30**, 460-465.

30. Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K. and Yeh, J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol*, **29**, 697-698.

31. Sung, Y.H., Baek, I.J., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.S. and Lee, H.W. (2013) Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol*, **31**, 23-24.

32. Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X., Zhang, L. *et al.* (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol*, **29**, 695-696.

33. Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y. *et al.* (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114-118.

34. Carlson, D.F., Tan, W., Lillico, S.G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D.F., Long, C.R., Whitelaw, C.B. and Fahrenkrug, S.C. (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A*, **109**, 17382-17387.

35. Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C. *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*, **29**, 731-734.

36. Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.

37. Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T. *et al.* (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell*, **12**, 238-251.

38. Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B. *et al.* (2012) Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *J Genet Genomics*, **39**, 209-215.

39. Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*, **27**, 851-857.

40. Osborn, M.J., Starker, C.G., McElroy, A.N., Webber, B.R., Riddle, M.J., Xia, L., Defeo, A.P., Gabriel, R., Schmidt, M., Von Kalle, C. *et al.* (2013) TALEN-based Gene Correction for Epidermolysis Bullosa. *Mol Ther*, doi: 10.1038/mt.2013.1056.

41. Meier, I.D., Bernreuther, C., Tilling, T., Neidhardt, J., Wong, Y.W., Schulze, C., Streichert, T. and Schachner, M. (2010) Short DNA sequences inserted for gene

targeting can accidentally interfere with off-target gene expression. *Faseb J*, **24**, 1714-1724.

42. Kaufman, D.S. (2009) Toward clinical therapies using hematopoietic cells derived from human pluripotent stem cells. *Blood*, **114**, 3513-3523.

43. Lengerke, C. and Daley, G.Q. (2010) Autologous blood cell therapies from pluripotent stem cells. *Blood Rev*, **24**, 27-37.

44. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.

45. Yusa, K., Zhou, L., Li, M.A., Bradley, A. and Craig, N.L. (2011) A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A*, **108**, 1531-1536.

CHAPTER 4. DIRECTED EVOLUTION OF TALENS WITH IMPROVED GENOME EDITING EFFICACY

4.1. Introduction

Genetic modifications of human genomes have a wide range of applications such as investigating human biology, studying disease mechanisms and treating diseases via gene therapies (1,2). Although virus-based transgenesis can modify human genomes efficiently by introduction of exogenous genes, random integration of viral transgenes poses the risk of insertional mutagenesis and carcinogenesis (3,4). In contrast, targeted genome editing or engineering enables researchers to tailor the human genome in a precise manner. To achieve efficient genome editing in human cells, DNA double-strand breaks (DSBs) have to be introduced at the specific sites of the chromosome by custom-designed DNA endonucleases. Subsequent repair of the DSBs by non-homologous end joining (NHEJ) or homologous recombination can generate desired genomic modifications such as gene disruptions, gene deletions, gene insertions, gene replacements and chromosome rearrangements (see **Chapter 1**).

Recently, transcription activator-like effector (TALE) nucleases (TALENs) have emerged as an efficient and versatile tool for genome editing by introduction of chromosomal DNA DSBs (5-7). TALENs are artificial DNA endonucleases composed of a TALE central repeat domain as the DNA recognition module and a non-specific FokI nuclease domain as the DNA cleavage module. The TALE central repeat domain comprises a series of tandem repeat units, each of which typically contains 34 residues.

Each repeat is used to recognize a single nucleotide and the DNA recognition specificity is determined by the highly variable residues at positions 12 and 13 (*e.g.* NI recognizes adenine, HD recognizes cytosine, NG recognizes thymine, NH recognizes guanine and NN recognizes guanine and adenine) (8-11). The simple DNA recognition code and the modular nature of the TALE central repeat units enable researchers to tailor DNA recognition specificity with ease and target essentially any desired DNA sequence. Therefore, TALENs have been widely applied for genome editing in various species (7). Due to their long recognition sequences (30-50 bp) and high DNA recognition specificity, TALENs exhibit minimal off-target effects in the context of complex human genomes (12-14), making them an ideal platform for human genome editing.

First-generation TALEN scaffolds have been described with different N-terminal segments (NTSs) and C-terminal segments (CTSs) flanking the TALE central repeat domains (13-16). Some of them have been applied for generating human stem cell-based disease models (12) and treating human diseases (17,18), but their efficacy of modifying human genomes is limited in targeting certain loci (16,19). Specifically, only ~5% of cells can be converted into GFP-positive in the human cell reporter assay as shown in **Chapter 2** and only 1 copy of the *HBB*^{S} gene can be corrected as shown in **Chapter 3**. Those evidences indicate a high-efficiency TALEN system in human cells is highly desirable for both scientific and therapeutic applications. Although a second-generation TALEN platform called GoldyTALEN has been reported with improved genome editing efficacy in zebrafish (20) and livestock (21), its efficacy in modifying human genomes has not been demonstrated. In this chapter, we report the development and application of a high-throughput screening system to improve TALEN activity through a directed

Table 4.1. Recognition sequences of TALENs. The TALE binding sites are highlighted.

Target site	Recognition sequence	Spacer length (bp)
AvrXa10	TATATAAGCACGTATCTTCTAGGATAGGACTAAAGATACGTGCTTATATA	16
HBB^{S}	TAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTA	15
ABL1	TACCTATTATTACTTTATGGGGCAGCAGCCTGGAAAAGTACTTGGGGACCAA	17
APC	TATGTACGCCTCCCTGGGCTCGGGTCCGGTCGCCCCTTTGCCCGCTTCTGTA	16
BRCA2	TTAGACTTAGGTAAGTAATGCAATATGGTAGACTGGGGAGAACTACAAACTA	16
CCND1	TGGAACACCAGCTCCTGTGCTGCGAAGTGGAAACCATCCGCCGCGCGTACCCCGA	16
CCR5	TGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACA	15
ERCC2	TCCGGCCGGCGCCATGAAGTGAGAAGGGGGCTGGGGGTCGCGCTCGCTA	16
FH	TGTACCGAGCACTTCGGCTCCTCGCGCGCTCGCGTCCCCTCGTGCGGGCTCCA	17
MYC	TGCTTAGACGCTGGATTTTTTTCGGGTAGTGGAAAACCAGGTAAGCACCGAA	16
PTEN	TCCCAGACATGACAGCCATCATCAAAGAGATCGTTAGCAGAAACAAAAGGA	16
SDHC	TGTTGCTGAGGTGACTTCAGTGGGACTGGGAGTTGGTGCCTGCGGCCCTCCGGA	16
SS18	TGGTGACGGCGGCAACATGTCTGTGGCTTTCGCGGCCCCGAGGCAGCGAGGCA	17
XPA	TGGGCCAGAGATGGCGGCGGCCGACGGGGCTTTGCCGGAGGCGGCGGCTTTA	16

evolution strategy. We successfully isolated a TALEN variant, SunnyTALEN, with significantly improved efficacy in modifying human genomes. We demonstrate that the corresponding TALEN scaffold is more active than the GoldyTALEN system in human cells and compatible with previously published obligate heterodimeric FokI nuclease domains. This novel second-generation TALEN system provides a general solution for efficient modifications of human genomes with low toxicities and has great potential in both basic and applied biological sciences.

4.2. Results

4.2.1. Development of a high-throughput screening system

To improve the performance of the existing TALEN technology using directed evolution, we developed a high throughput screening system in *Saccharomyces cerevisiae*. This system employs an enhanced green fluorescent protein (eGFP)-based single-strand annealing reporter construct that is integrated into the yeast genome. The reporter construct contains a divided *eGFP* gene harboring two AvrXa10 TALE binding sites in a

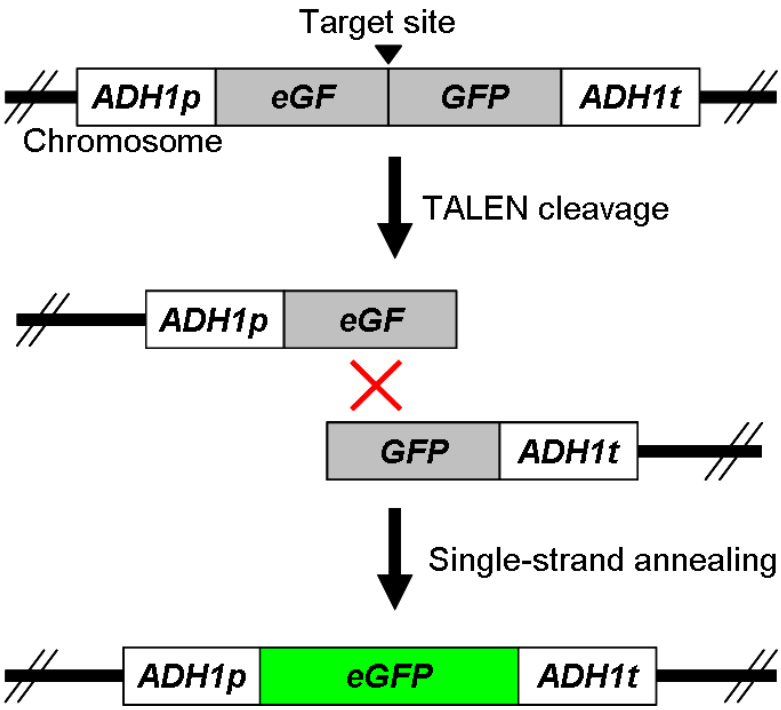

Figure 4.1. Schematic of an eGFP-based single-strand annealing reporter system in yeast. *ADH1p*, promoter of the *ADH1* gene; *ADH1t*, transcription terminator of the *ADH1* gene; *eGF*, 5'-fragment of the *eGFP* gene; and *GFP*, 3'-fragment of the *eGFP* gene.

tail-to-tail orientation (**Figure 2.1C** and **Table 4.1**). The DSB generated by the AvrXa10 TALEN homodimer can induce single-strand annealing of the two truncated *eGFP* DNA fragments and reconstitute a complete and functional *eGFP* gene (**Figure 4.1**). Therefore, this system couples the enzymatic DNA cleavage activity with the fluorescent signal of host cells, and enables high-throughput screening of TALEN mutants with improved genome editing efficacy. The TALEN gene expression is under the control of a galactose-inducible promoter, thus the selection pressure of directed evolution can be adjusted by galactose concentration in the growth medium and incubation time during induction (**Figure 4.2**).

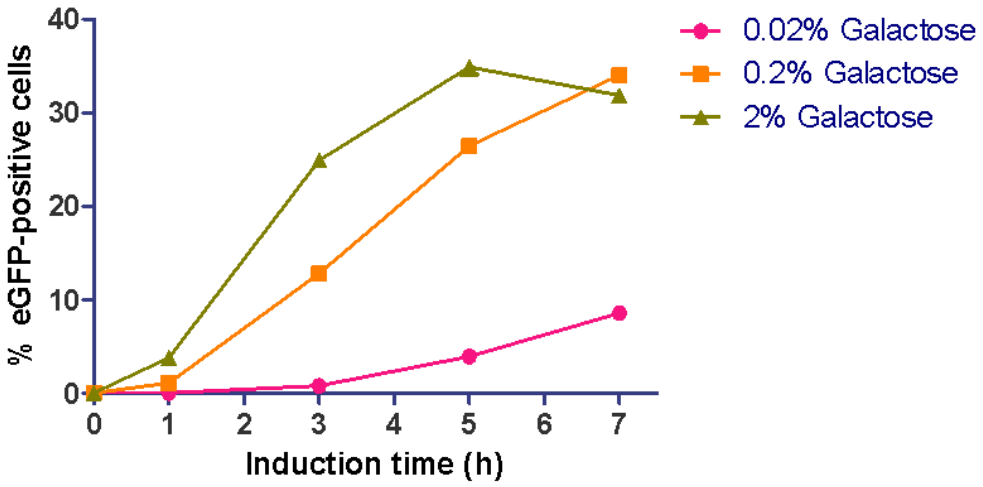

Figure 4.2. Activities of the WT TALEN in the eGFP-based yeast screening system under different induction conditions.

4.2.2. Directed evolution of TALENs with improved activity

Using a TALEN scaffold bearing a 207 aa NTS and a 63 aa CTS (hereinafter referred to as WT) (14) as a template, we randomly mutagenized the sequences encoding the CTS and the FokI nuclease domain using error-prone PCR and constructed a library of mutants *in vivo* by gap repair homologous recombination in yeast (**Figure 4.3**). TALEN expression was induced by galactose and the eGFP-positive cells were collected by fluorescence-activated cell sorting (FACS). The sorted cells were pooled together and plasmids from the pool were isolated. The regions encoding the CTS and FokI nuclease domain of TALEN genes were PCR-amplified from the plasmids and subjected to another round of screening. This process was continued in an iterative fashion until the TALEN mutants with improved genome editing activity were enriched and identified (**Figure 4.4**). We gradually increased the selection stringency by decreasing the galactose concentration and induction time during the enrichment (**Figure 4.5**). After two rounds of

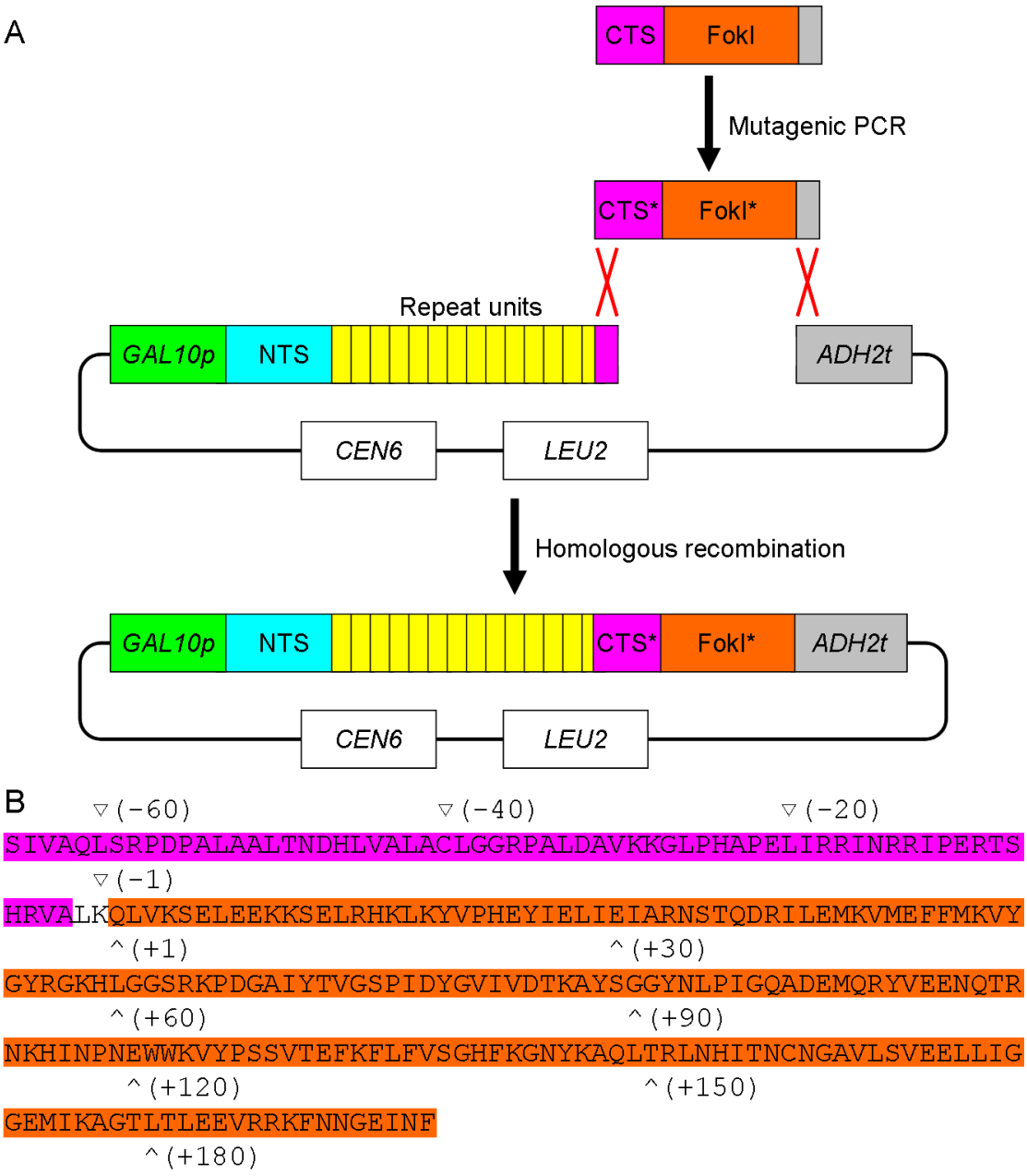

Figure 4.3. Strategy for the creation of a TALEN mutant library. (**A**) Schematic depicting random mutagenesis of the CTS and the FokI nuclease domain and assembly of the library of mutants through homologous recombination in yeast. *GAL10p*, promoter of *GAL10* gene; *ADH2t*, transcription terminator of *ADH2* gene; and *CEN6*, a centromere of yeast chromosome. *LEU2* auxotrophic marker is used for selection of transformants in yeast. Asterisks indicate mutagenized CTS and FokI. (**B**) Sequence and numbering of the CTS and the FokI domain. The CTS is highlighted in magenta. The FokI domain is highlighted in orange.

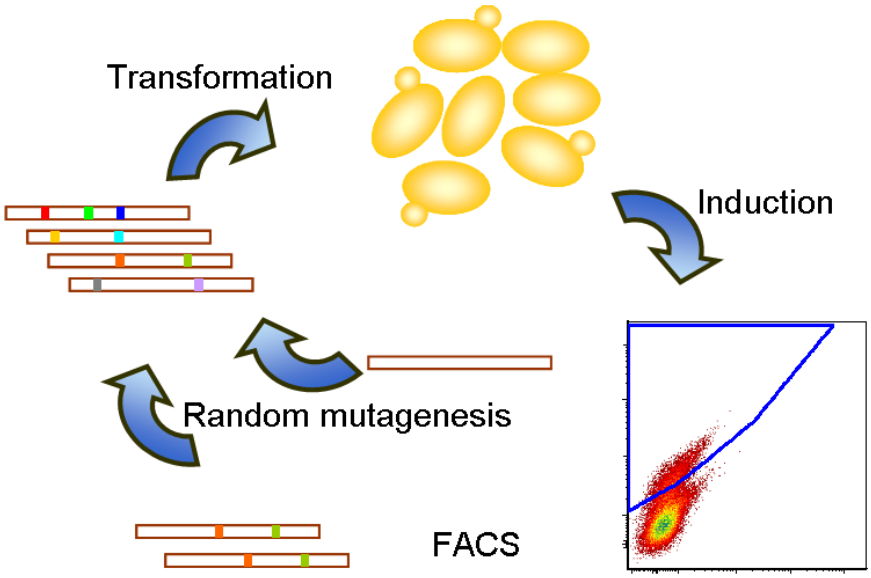

Figure 4.4. Schematic representation of the directed evolution approach in yeast for isolating novel TALEN variants.

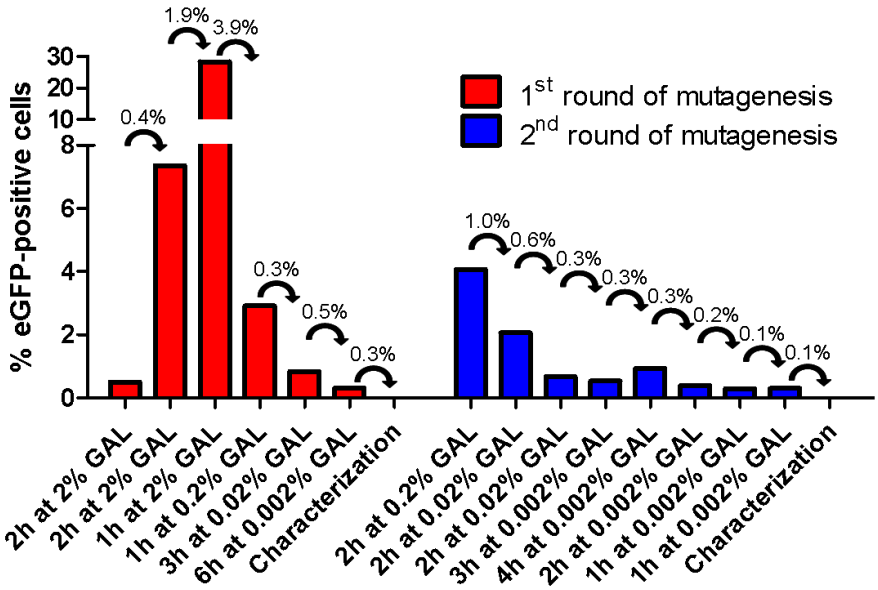

Figure 4.5. Directed evolution of TALENs in the yeast screening system. The columns denote the percentage of eGFP-positive cells in each screening cycle and the induction of TALEN expression was performed in galactose (GAL)-containing medium for certain amounts of time. The arrows indicate the portions of cells that were collected by FACS and applied for the next cycle of screening.

Table 4.2. Nomenclature, sequence information and *in vivo* activities of TALEN variants. The residue numbering is indicated in **Figure 4.3B**. Silent mutations are shown in grey. Activities in yeast were measured when the TALEN expression was induced in the cultural medium containing 0.02% galactose for 4 h. Activities in human cells were measured in the modified surrogate reporter system (**Figure 4.6**) in a dose-limiting condition (32 ng transfected TALEN expression plasmids per 10^5 cells). The activity of WT was normalized to 1. Values represent the average (± standard deviation) of at least three independent experiments.

Label	Mutation	Relative activity (fold of WT)	
		yeast	human cells
WT	-	1.00 ± 0.11	1.00 ± 0.05
Sharkey	S35P, K58E	2.38 ± 0.23	1.01 ± 0.02
Goldy[a]	L(-33)M, I(-16)V, P(-11)G	3.24 ± 0.15	0.80 ± 0.06
S4C5	P(-11)T, H(-6)Q, V52I	3.49 ± 0.14	2.08 ± 0.10
S4C9	P(-11)H, L182L, R186R	2.90 ± 0.32	2.26 ± 0.11
S4C10	K58R, N144K	1.30 ± 0.10	0.40 ± 0.03
S7C1	P(-11)R, I82I, H115H, N153K, L168L, L182L	4.40 ± 0.16	1.11 ± 0.03
S7C6	P(-11)H	4.11 ± 0.23	2.26 ± 0.03
S7C7	S5S, I175I, K176Q, L182L	2.83 ± 0.17	1.62 ± 0.08
S7C10	K58E, I70I, L182L, E193V	1.76 ± 0.22	0.10 ± 0.01
Sunny	P(-11)H, S35P, K58E	4.13 ± 0.27	2.69 ± 0.43
2S3C3	P(-11)H, S35P, K51E, K58E	7.26 ± 0.38	2.21 ± 0.14
2S3C11	P(-11)H, S35P, K58E, A87A, F134Y	6.85 ± 0.42	0.91 ± 0.02
2S3C12	L(-33)L, P(-11)H, L18L, S35P, K58E, K123N	4.91 ± 0.40	1.96 ± 0.09
2S3C18	P(-11)H, E(-10)K, L7L, S35P, K58E, I70M, P76L, V81V	6.20 ± 0.57	1.69 ± 0.14
2S6C1	P(-11)H, N34D, K58E	5.88 ± 0.56	1.10 ± 0.05
2S8C6	P(-11)H, E(-10)K, S35P, K58E	5.67 ± 0.47	2.03 ± 0.19

[a] GoldyTALEN has a shorter and mutated NTS compared with all the other TALENs.

mutagenesis and evolution, fourteen TALEN mutants with improved activity were generated (**Table 4.2**). The most active mutant (2S3C3) has greater than sevenfold improvement in yeast genome modification compared to WT, whereas GoldyTALEN and the TALEN bearing a Sharkey domain (a FokI domain variant with S35P and K58E mutations) (22) only exhibit <3.5-fold increase compared to WT.

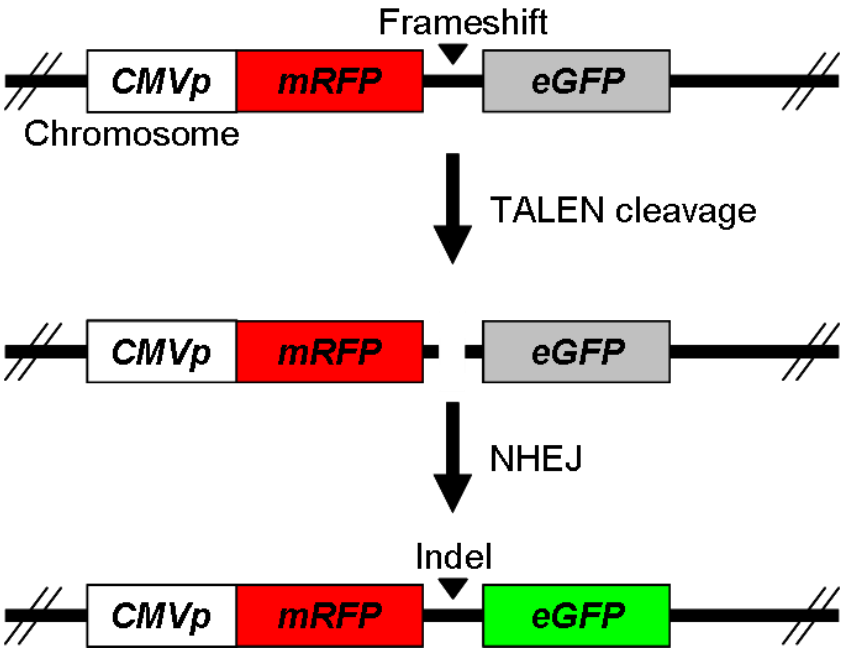

Figure 4.6. Schematic of a modified surrogate reporter system in Hela cells. *CMVp*, *CMV* promoter.

4.2.3. Isolation and characterization of the SunnyTALEN scaffold

To test the efficacy of the isolated TALEN mutants in human cells, we modified a surrogate reporter system (23) to rapidly and quantitatively gauge the potential of TALEN-driven genome editing (**Figure 4.6**). The reporter construct encoding a monomeric red fluorescent protein (mRFP)-eGFP fusion protein was stably integrated into the genome of Hela cells. A TALEN target site within sickle human β-globin (HBB^S) gene locus (**Table 4.1**) (14) was inserted between the *mRFP* and *eGFP* genes, making the *eGFP* gene out of frame. In response to the TALEN-induced DSBs, certain insertions and deletions (indels) caused by NHEJ-mediated mutagenesis will make the *eGFP* gene in frame and restore eGFP function. Each TALEN mutant was under the control of a cytomegalovirus (*CMV*) promoter, and genome editing activity was measured by flow cytometry following transient expression. Although the TALEN bearing a Sharkey

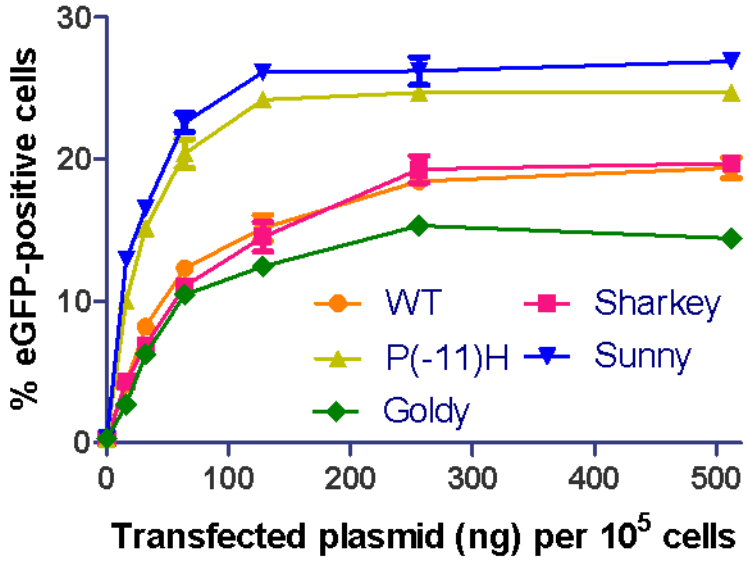

Figure 4.7. Dose-response curve using titrations of different TALEN expression plasmids as tested in the modified surrogate reporter system in Hela cells. Error bars indicate standard deviation of three independent experiments.

domain and GoldyTALEN showed similar or lower genome editing efficacy compared to WT, one TALEN mutant (referred to as SunnyTALEN) has >2.5-fold improvement compared to WT (**Table 4.2**). SunnyTALEN contains a single P(-11)H mutation on the CTS and a Sharkey domain. We performed titrations of the expression plasmids of different TALEN architectures at various amounts and obtained a dose-response curve (**Figure 4.7**). SunnyTALEN consistently yielded more eGFP-positive cells compared to the other TALEN variants. The P(-11)H mutation of SunnyTALEN serves as a major contributor to its improved efficacy without changing protein expression level. The S35P and K58E mutations on the Sharkey domain increase activity marginally, which can be explained by the increased protein solubility (**Figure 4.8**). The lower protein expression level of GoldyTALEN is consistent with its lower efficacy in human cells than WT.

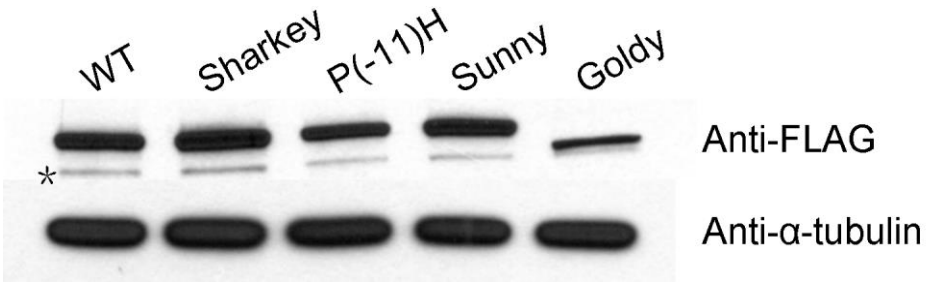

Figure 4.8. Western blot analysis of TALEN variants in human cells. All the TALEN variants contained a FLAG tag at the N-terminus. α-Tubulin was used as a loading control. An asterisk indicates non-specific bands.

4.2.4. Application of the SunnyTALEN scaffold for human genome editing

To further explore the general applicability of the SunnyTALEN scaffold in modifying human genomes, we generated TALENs against 12 additional genomic loci of human embryonic kidney (HEK) 293 cells. Using the same TALE DNA binding domains in the SunnyTALEN scaffold, we observed increased rate of gene modification at all the targeted loci compared to WT, with the increased indels ranging from 4% to 33% in response to NHEJ-mediated mutagenesis (**Figure 4.9**). Sequencing analysis of the two targeted loci indicated that SunnyTALENs induced more small deletions than insertions (**Figure 4.10**), with the mutation signatures similar to the first-generation TALENs (24) and GoldyTALENs (20). We also explored whether the rates of TALEN-induced homologous recombination could be enhanced using an *eGFP* gene conversion assay (**Figure 2.5** and **Figure 4.11A**) (14). In the linear range of the dose-response curve, we observed a >2-fold increase in homology-directed gene repair using the SunnyTALEN scaffold over WT (**Figure 4.11B**), which paralleled the increase observed in NHEJ-mediated gene disruption levels (**Figure 4.7**).

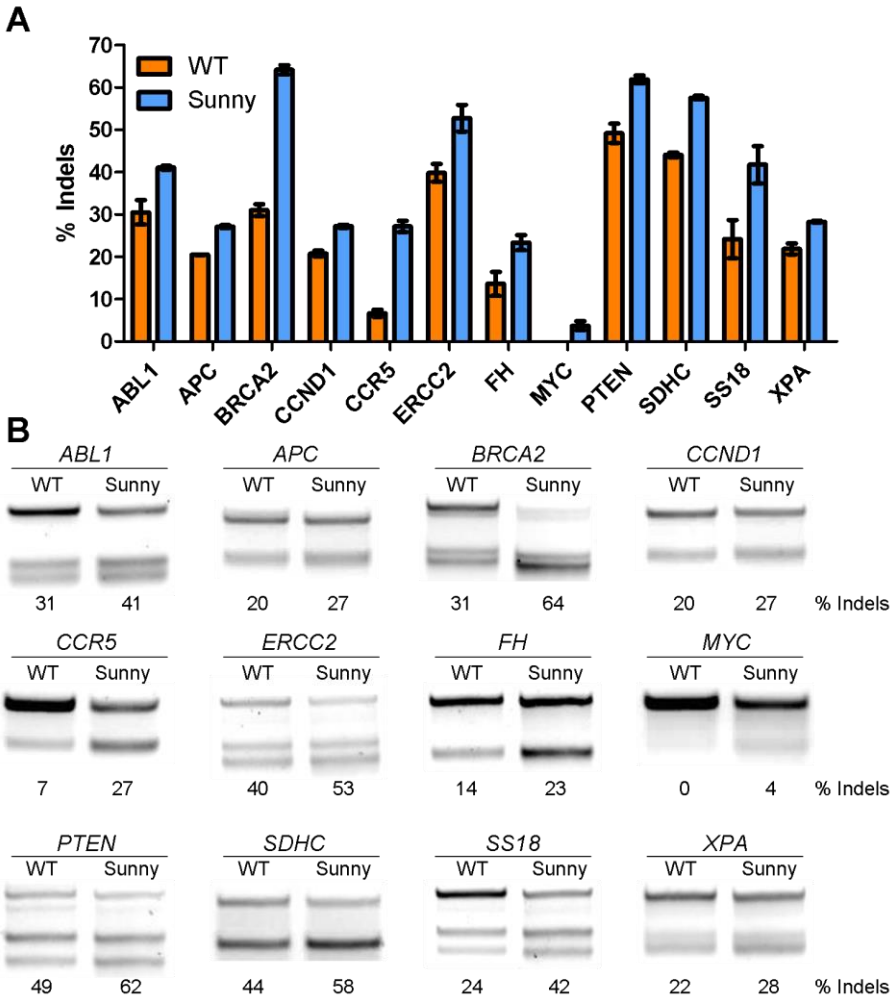

Figure 4.9. Application of the SunnyTALEN scaffold for human genome editing. (**A**)The SunnyTALEN scaffold improved NHEJ-mediated gene disruptions at 12 genomic loci in HEK293 cells. (**B**) Representative gel images of the SURVEYOR nuclease assay to determine the frequencies of TALEN-induced indels.

To prevent the formation of unexpected cleavage-competent homodimers, obligate heterodimeric FokI domains have been used in TALENs to reduce off-target effects and cytotoxicities (25,26). To investigate the compatibility of the SunnyTALEN scaffold with the obligate heterodimeric FokI domains, we introduced the ELD:KKR FokI pair (ELD denotes Q103E, I116L and N113D mutations; KKR denotes E107K, I155K and H154R

BRCA2

5′-TTAGACTTAGGTAAGTAATGCAATATGGTAGACTGGGGAGAACTACAAACTA-3′

Mutated sequences

5′-TTAGACTTAGGTAAGTAATGCAAT----TAGACTGGGGAGAACTACAAACTA-3′
5′-TTAGACTTAGGTAAGTAATGCAAT--GGTAGACTGGGGAGAACTACAAACTA-3′
5′-TTAGACTTAGGTAAGTAATGCAA---------CTGGGGAGAACTACAAACTA-3′
5′-TTAGACTTAGGTAAGTAATGCAAT---G-AGACTGGGGAGAACTACAAACTA-3′
5′-TTAGACTTAGGTAAGTAATGCAATA---TAGACTGGGGAGAACTACAAACTA-3′
5′-TTAGACTTAGGTAAGTAATGCAATA-----GACTGGGGAGAACTACAAACTA-3′
5′-TTAGACTTAGGTAAGTAATGCAAT---------TGGGGAGAACTACAAACTA-3′

SS18

5′-TGGTGACGGCGGCAACATGTCTGTGGCTTTCGCGGCCCCGAGGCAGCGAGGCAAG-3′

Mutated sequences

5′-TGGTGACGGCGGCAA--G-3′
5′-TGGTGACGGCGGCAA----------G--TTCGCGGCCCCGAGGCAGCGAGGCAAG-3′
5′-TGGTGACGGCGGCAAC----------CTTTCGCGGCCCCGAGGCAGCGAGGCAAG-3′
5′-TGGTGACGGCGGCAACATGTCTGTG---TT-------CCGAGGCAGCGAGGCAAG-3′
5′-TGGTGACGGCGGC----------------------CCCGAGGCAGCGAGGCAAG-3′
5′-TGGTGACGGCGGCAACATGTCTGTGGGGCTGCTGCTTTCGCGGCCCCGAGGCAGCGAGGCAAG-3′

Figure 4.10. Sequencing analysis of the mutagenesis generated by SunnyTALEN-induced NHEJ against *BRCA2* and *SS18* sites in HEK293 cells. TALEN binding sites are highlighted in yellow. Deleted nucleotides are shown in dashes. Inserted nucleotides are shown in red.

mutations) (27) that preferentially heterodimerize into the SunnyTALEN scaffold and compared the genome editing efficacy with the WT counterpart in the modified surrogate reporter system (**Figure 4.6**). The pairwise analysis of these architectures showed that the TALEN heterodimers with SunnyTALEN scaffold had >4.4-fold improvement at stimulating mutagenesis compared to the WT TALEN heterodimers (**Figure 4.12**). None of the ELD:ELD or KKR:KKR TALEN dimers were observed to be cleavage-competent. To assess the nuclease-associated toxicity, we monitored the phosphorylation of histone H2AX (γ-H2AX), which rapidly responds to chromosomal DSBs and can be used to measure genome-wide DNA damage levels. Although the SunnyTALEN scaffold slightly

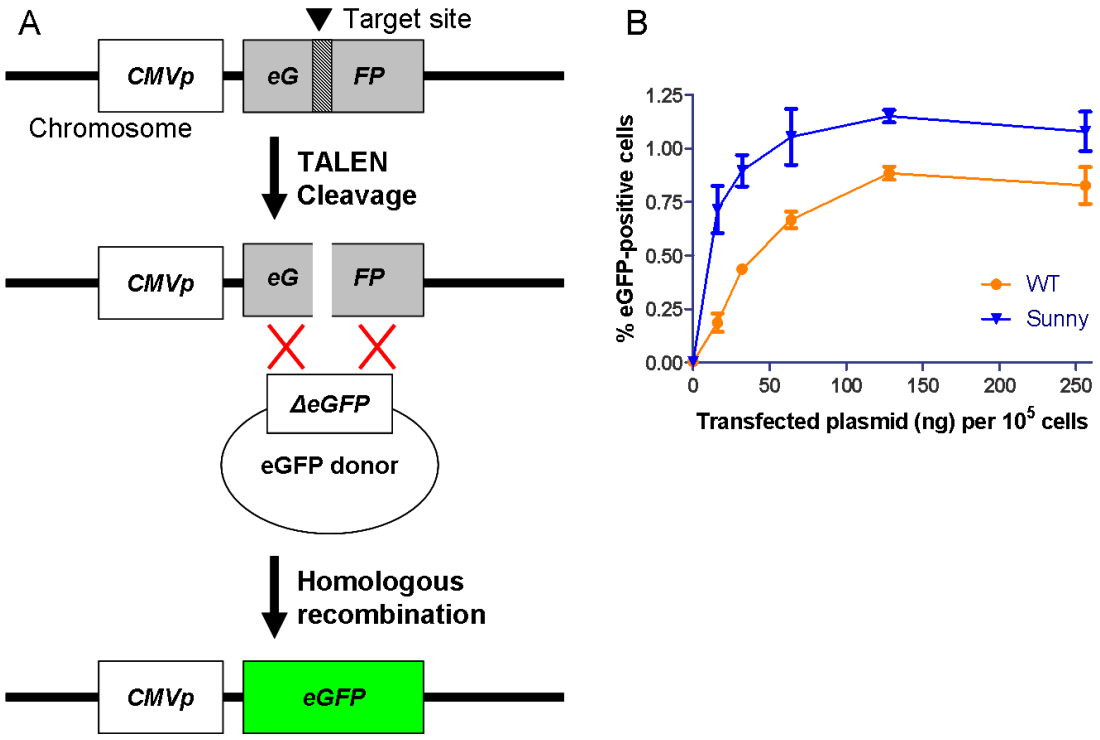

Figure 4.11. The SunnyTALEN platform enhanced the efficiency of homologous recombination in human cells. (**A**) Schematic of an *eGFP* gene conversion assay in Hela cells. (**B**) Dose-response curve using titrations of different TALEN expression plasmids as tested in the *eGFP* gene conversion assay. Error bars indicate standard deviation of three independent experiments.

increased γ-H2AX level over WT, the introduction of the obligate heterodimeric FokI domains reduced γ-H2AX staining to the background level (**Figure 4.13**).

4.3. Discussion

In this study, we developed a powerful yeast-based screening system to improve TALEN activity. This system couples enzymatic DNA cleavage with GFP signal of yeast cells and enables directed evolution of TALENs with improved genome editing efficacy. The GFP-positive cells can be isolated by FACS, which enables high-throughput screening of

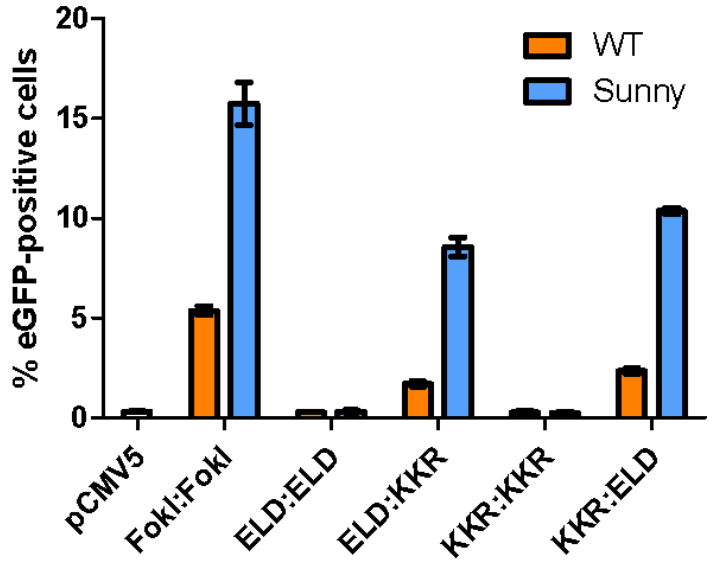

Figure 4.12. The SunnyTALEN scaffold is compatible with the obligate heterodimeric FokI nuclease domains. Activities were measured in the modified surrogate reporter system (**Figure 4.6**) in a dose-limiting condition (32 ng transfected TALEN expression plasmids per 10^5 cells). Error bars indicate standard deviation of three independent experiments. Empty vector pCMV5 served as a negative control.

large libraries of mutants with single cell resolution and low false positive rate. Although applied here in enhancing TALEN activity, this system can be easily adapted to engineer other rare-cutting DNA endonucleases such as homing endonucleases (or meganucleases) (28), zinc finger nucleases (ZFNs) (29), and newly developed clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system (30-34), for the purpose of increasing catalytic activities or altering DNA recognition specificities.

In the SunnyTALEN scaffold, the P(-11)H mutation on the CTS plays a major role in activity improvement. Because structural information is not available to visualize that residue, we have built structural models of WT and SunnyTALEN scaffolds with 63 aa

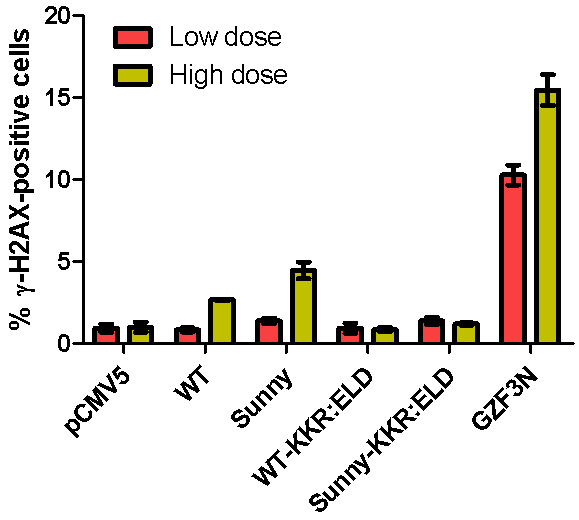

Figure 4.13. Nuclease-associated cytotoxicities of TALENs. Low dose indicates 32 ng TALEN expression plasmids were transfected per 10^5 cells. High dose indicates 256 ng TALEN expression plasmids were transfected per 10^5 cells. A reported toxic zinc finger nuclease (GZF3N) served as a positive control (35). Error bars indicate standard deviation of three independent experiments.

CTSs to study the molecular mechanism. Structural models indicate the substitution of a histidine for a proline introduces an additional hydrogen bond between the CTS of the SunnyTALEN scaffold and the DNA backbone, which might increase catalytic activity by positioning the FokI nuclease domain closer to the DNA substrate or to the FokI domain of the other TALEN monomer (**Figure 4.14**). More detailed structural studies will be needed to confirm this hypothesis. In addition to the isolation of a high-efficiency TALEN system, our directed evolution endeavor identified "hot spot" residues (*e.g.* P(-11), E(-10), N34, S35, K51, K58, F134 and K176) that are critical for the catalytic activity or protein solubility of TALENs (**Table 4.2**), which provides an important insight for the further improvement of TALEN system by rational protein design or saturation mutagenesis.

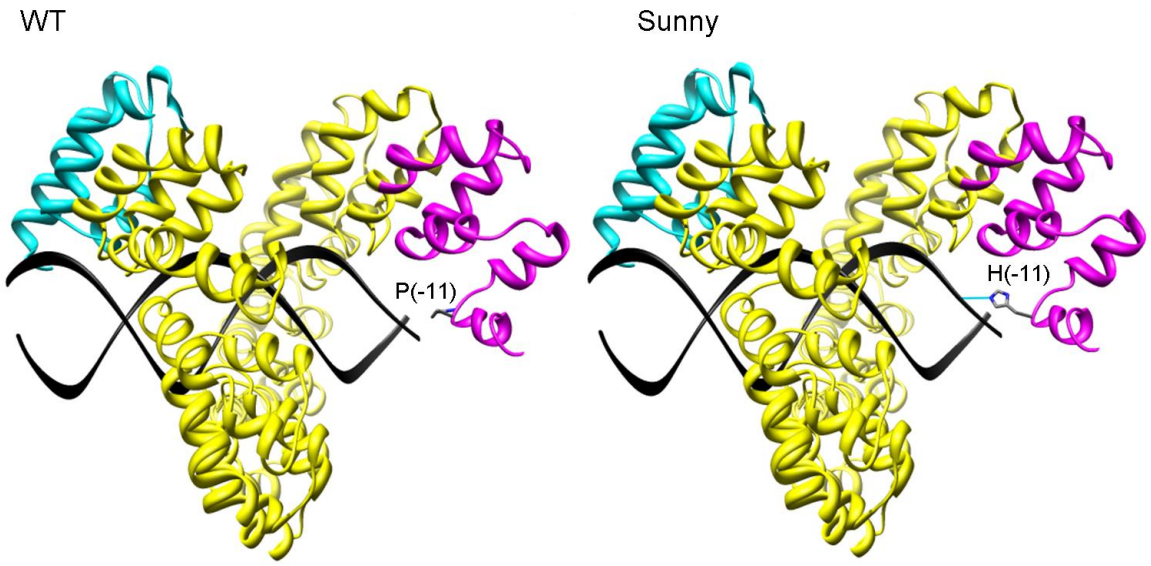

Figure 4.14. Structural models of the WT and SunnyTALEN scaffolds. DNA substrates are shown in black ribbons. The NTSs are shown in cyan. The central repeat units are shown in yellow. The CTSs are shown in magenta. A hydrogen bond is shown in cyan.

Using γ-H2AX as a marker of chromosomal DNA damage, we measured the nuclease-associated genotoxicities in human cells. Even though the increased genome editing efficacy of the SunnyTALEN scaffold is accompanied by increased DNA damage, γ-H2AX level can be reduced by lowering the TALEN dose or introducing obligate heterodimeric FokI nuclease domains (**Figure 4.13**). Because of the high efficiency of the SunnyTALEN scaffold, it can generate sufficient genetic modifications at dose-limiting conditions without causing detectable genotoxicities (**Figure 4.7** and **Figure 4.13**). Because off-target cleavage is dependent on protein concentration as demonstrated for ZFNs (36), it is suggested to optimize the dose of SunnyTALENs for different human cell lines to achieve efficient genome editing with low toxicity. To avoid the off-target events generated by homodimeric ZFNs, obligate heterodimer mutations were introduced at the dimer interface of the FokI cleavage domain to prevent homodimerization based on

electrostatic and hydrophobic interactions (27,35,37). Similar principle was applied to TALENs and successfully reduced off-target cleavage and relieved toxicity (25). We show here that the incorporation of obligate heterodimeric FokI domains completely abolished the formation of cleavage-competent SunnyTALEN homodimers, resulting in reduced genotoxicity (**Figure 4.12** and **Figure 4.13**).

In conclusion, we isolated the high-efficiency SunnyTALEN scaffold using a directed evolution strategy. The SunnyTALEN scaffold is portable to many TALE DNA binding domains and compatible with heterodimer TALEN architectures. Although deployed here in yeast and human cells, the SunnyTALEN scaffold could be effective in various organisms and cell types. In addition to the GoldyTALEN scaffold that enables efficient genome modifications in zebrafish and livestock, the SunnyTALEN scaffold serves as a novel second-generation TALEN system and will aid in a variety of applications in medicine and biology.

4.4. Materials and Methods

4.4.1. Materials

Q5 DNA polymerase, T4 DNA ligase, Antarctic phosphatase, and restriction endonucleases were purchased from New England Biolabs (Beverly, MA). QIAprep Spin Plasmid Miniprep Kit, QIAquick Gel Extraction Kit, and QIAquick PCR Purification Kit were obtained from Qiagen (Valencia, CA). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). All the other reagents unless specified were obtained from Sigma-Aldrich (St. Louis, MO).

4.4.2. Yeast reporter strain

To construct a single-strand annealing eGFP reporter, two separated *eGFP* fragments sharing a 100 bp homologous region were PCR amplified separately, digested with *Avr*II and *Xba*I, respectively, and ligated overnight at 16 °C. The ligation product was subsequently digested with a combination of *Avr*II and *Xba*I and the 870 bp non-cut fragment was gel-purified. The *eGFP* reporter gene was then cloned into the pRS414 plasmid (New England Biolabs) between the *ADH1* promoter and the *ADH1* terminator. The centromere sequence of the plasmid was then removed by the *Pml*I digestion and blunt ligation. This plasmid was then digested with *Hind*III at the *TRP1* gene. The linearized plasmid was transformed into *S. cerevisiae* HZ848 (*MATα, ade*2-1, *ade*3Δ22, *Δ*ur*a*3, *his*3-11,15, *trp*1-1, *leu*2-3,112 and *can*1-100) using the LiAc/SS carrier DNA/PEG method (38) and stably integrated into the chromosome at the *TRP1* site. The yeast reporter strain was obtained by selection of the transformed yeast cells on plates containing synthetic complete medium lacking tryptophan with 2% glucose at 30 °C. The integration was confirmed by PCR amplification and DNA sequencing of the targeted genomic region.

4.4.3. Library creation

The pRS415 yeast expression vector encoding the AvrXa10 TALEN that has been described previously (14) was used as the template for mutagenic PCR. The sequence encoding the CTS and FokI nuclease domain was mutagenized using the GeneMorph II Random Mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer's instruction. The following primers were used for the amplification: EP-FokI-Far-for 5’-

ttg ttg ccc agt tat ctc gc-3' and EP-FokI-Far-rev 5'-cgt gaa act tcg aac act gtc-3'. Two libraries with a low mutation rate (0-4.5 mutations/kbp) and a medium mutation rate (4.5-9 mutations/kbp) were amplified and gel-purified, respectively. The TALEN expression plasmid was digested with *Aat*II and *Sal*I and gel-purified to remove the sequences encoding the wild-type CTS and FokI domain. The two gene libraries were mixed and co-transformed with the linearized TALEN expression plasmids at an approximately 13:1 insert:vector molar ratio into the yeast reporter strain using the LiAc/SS carrier DNA/PEG method (38). Through yeast homologous recombination, the pRS415 plasmid library was created containing ~2×10^5 transformants.

4.4.4. High-throughput screening

After transformation, the yeast reporter strain carrying the library of TALEN variants was recovered at 30 °C with shaking for 1 h in YPA medium (1% yeast extract, 2% peptone and 0.01% adenine hemisulphate) with 2% glucose. The cells were then centrifuged and resuspended in synthetic complete medium lacking leucine and tryptophan (SC-Leu-Trp) with 2% raffinose for growing at 30 °C with shaking overnight. TALEN expression was induced by culturing cells in YPA medium with 0.002-2% galactose at 30 °C for 1-6 h (the induction conditions for each cycle of screening are listed in **Figure 4.5**). The cells were then centrifuged and resuspended in SC-Leu-Trp liquid medium with 2% glucose at 30 °C with shaking overnight. On the next day, cells were analyzed on a BD FACS Aria III cell sorting system (BD Biosciences, San Jose, CA) and 20,000-100,000 eGFP-positive cells were collected. After 2 days' growth at 30 °C in SC-Leu-Trp liquid medium with 2% glucose, plasmids from the sorted cells were extracted using Zymoprep Yeast

Plasmid Miniprep II Kit (Zymo Research, Orange, CA). Sequences encoding the CTS and FokI domain were PCR amplified from the plasmids by Q5 DNA polymerase and gel-purified for the next cycle of screening.

4.4.5. A modified surrogate reporter system

A TALEN target site within the *HBB*^{S} gene was inserted between the *mRFP* gene and the *eGFP* gene by overlap extension PCR. The fusion gene was cloned into the pLNCX2 retroviral vector (BD Clontech, Palo Alto, CA) in the presence of the *CMV* promoter and stably integrated into the genome of Hela cells through retroviral transduction according to the manufacturer's protocol. Transfected cells were selected in 500 μg/mL G418 for two weeks. The TALEN genes along with a FLAG tag sequence and a SV40 nucleus localization signal added to the N-terminus were cloned into the pCMV5 mammalian expression vector (39) through the *Kpn*I and *Sal*I sites.

Reporter Hela cells were routinely maintained in the modified Eagle's medium (MEM) supplemented with 10% Fetal Bovine Serum (FBS; Hyclone, Logan, UT). Cells were seeded in 24-well plates at a density of 5×10^{4} per well. After 24 h, reporter cells were transfected with certain amounts of TALEN expression plasmids using FuGene HD transfection reagent (Promega, Madison, WI) under conditions specified by the manufacturer. For each transfection, the overall plasmid amount was made to be 500 ng by adding the empty pCMV5 plasmid. Cells were trypsinized from their culturing plates 48 h after transfection and resuspended in 300 μL phosphate buffered saline (PBS) for flow cytometry analysis. 20 000 cells were analyzed by a BD LSRII flow cytometer (BD Biosciences, San Jose, CA) to quantify the eGFP-positive cells.

4.4.6. Western blot analysis

Hela cells in 12-well plates were harvested 24 h after transfection with pCMV5-TALEN-HBB-R plasmid (14). Cells were collected by centrifugation, washed with PBS and resuspended in 40 μL whole cell lysis buffer (1 M Tris-HCl, pH 6.8, 20% sodium dodecyl sulfate, and 0.1 M dithiothreitol). Proteins were resolved by 4-20% Mini-Protean TGX Precast Gel (Bio-Rad Laboratories, Hercules, CA), transferred onto a nitrocellulose membrane, blocked for 1 h with Tris-buffered saline/0.05% Tween 20 containing 5% nonfat milk, followed by incubation with anti-FLAG tag (1: 500) and anti-α-tubulin (1: 10 000) antibodies at 4 °C overnight. After incubation with anti-mouse horseradish peroxidase-conjugated secondary antibody (1: 25 000; GeneScript, Piscataway, NJ) for 1 h, bands were visualized using SuperSignal West Pico Chemiluminescent Substrate (Pierce, Rockford, IL).

4.4.7. TALEN construction

The central repeat units of TALEs were assembled by the Golden Gate reactions as previously described (40). The receiver plasmids pCMV5-GG-WT for WT TALEN and pCMV5-GG-Sunny for the SunnyTALEN scaffold are available upon request. The NTS and CTS of the GoldyTALEN scaffold were PCR-amplified from pC-GoldyTALEN plasmid (Kindly provided by Dr. Daniel Carlson from University of Minnesota, Minneapolis, MN).

4.4.8. SURVEYOR nuclease assay

HEK293 cells in 6-well plates were harvested 48 h after transfection. Genomic DNA was

extracted using Wizard Genomic DNA Purification Kit (Promega, Madison, WI) according to the manufacturer's instructions. DNA fragments containing TALEN target sites were PCR-amplified from the genomic DNA using the primers described previously (19). Gel-purified PCR products were re-annealed slowly and analyzed by SURVEYOR Mutation Detection Kits (Transgenomic, Omaha, NE) according to the manufacturer's instructions. The product was resolved on a 2% agarose gel stained with GelStar (Lonza, Rockland, ME). The bands were quantified by G Box using Gene Tools software (Syngene, Frederick, MD). The apparent indel percentage of the original cell pool was calculated using the following equations as previously described (41):

Fraction cleaved = volume of cleaved bands/(volume of cleaved bands + volume of uncleaved bands)

% Indels = $100 \times (1- (1- \text{Fraction cleaved})^{1/2})$

4.4.9. H2AX phosphorylation assay

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS. Cells were seeded in 12-well plates (2×10^5 per well) and transfected after 24 h with 500 ng of each TALEN expression plasmid, 1 μg empty pCMV5 vector as a negative control or 1 μg reported toxic ZFN construct GZF3N (kindly provided by Dr. Toni Cathomen of Hannover Medical School, Hannover, Germany) as a positive control. After 48 h post transfection, cells were harvested, fixed, permeabilized, and stained using the H2AX phosphorylation assay kit (Millipore, Watford, UK) according to the manufacturer's protocol. Cells were then scanned in a

flow cytometer to quantitate the number of cells staining positive for phosphorylated histone H2AX.

4.4.10. Sequencing analysis of endogenous gene mutations

HEK293 cells in 6-well plates were harvested 48 h after transfection with SunnyTALEN expression plasmids. Genomic DNA was extracted using Wizard Genomic DNA Purification Kit (Promega, Madison, WI) according to the manufacturer's instructions. DNA fragments containing TALEN target sites within the *BRCA2* gene were PCR-amplified from the genomic DNA using the primers KpnI-BRCA2-for 5'-act gac *ggt acc* tga tct tta act gtt ctg ggt cac aaa-3' (*Kpn*I recognition site shown in italics) and SalI-BRCA2-rev 5'-act gac *gtc gac* cgc cag gga aac tcc ttc ca-3' (*Sal*I recognition site shown in italics). DNA fragments containing TALEN target sites within the *SS18* gene were PCR-amplified from the genomic DNA using the primers KpnI-SS18-for 5'-act gac *ggt acc* ggg atg cag gga cgg tca ag-3' (*Kpn*I recognition site shown in italics) and SalI-SS18-rev 5'-act gac *gtc gac* gcc gcc cca tcc cta gag aaa-3' (*Sal*I recognition site shown in italics). PCR products corresponding to the two sites were cloned into pCMV5 plasmid between *Kpn*I and *Sal*I restriction sites. Plasmid DNA was isolated from 20 colonies for each transformation and then sent for Sanger sequencing using the primer pSeq-pCMV5-for 5'-cgc aaa tgg gcg gta ggc gtg-3'.

4.4.11. An *eGFP* gene conversion assay

The *eGFP* gene was divided into two fragments by insertion of a preselected sequence from the HBB^S gene locus and an in-frame stop codon. The non-functional *eGFP* reporter

stably integrated into the genome of Hela cells as previously described (14). A donor plasmid was constructed by insertion of a promoter-less *eGFP* gene with the first 37 nucleotides missing into the pNEB193 plasmid (New England Biolabs, Beverly, MA) as previously decribed (14). The donor plasmid provides a homologous DNA segment as a template for repairing the double-stranded break in the *eGFP* reporter. The reporter Hela cells were maintained in modified Eagle's medium (MEM) supplemented with 10% Fetal Bovine Serum (FBS; Hyclone, Logan, UT). Cells were seeded in 12-well plates at a density of 1×10^5 per well. After 24 hr, reporter cells were co-transfected with a certain amount of TALEN expression plasmids and 500 ng of donor plasmid using FuGene HD transfection reagent (Promega, Madison, WI) under conditions specified by the manufacturer. Cells were trypsinized from their culturing plates 48h after transfection and resuspended in 300 μL phosphate buffered saline (PBS) for flow cytometry analysis. 20,000 cells were analyzed by BD LSRII flow cytometer (BD Biosciences, San Jose, CA) to determine the percentage of eGFP-positive cells.

4.4.12. Computational modeling

The homology models of TALEs with 63 aa CTSs were created by the I-TASSER online server (42) based on the crystal structure of dHax3 (PDB accession code: 3V6P) (43). The DNA-bound models were created based on the crystal structure of DNA-bound dHax3 (PDB accession code: 3V6T) (43) and energy minimized using the Molecular Operating Environment (MOE, The Chemical Computing Group, 2008) software package. The figures were generated using UCSF Chimera (44).

4.5. References

1. Dow, L.E. and Lowe, S.W. (2012) Life in the fast lane: mammalian disease models in the genomics era. *Cell*, **148**, 1099-1109.
2. Cherry, A.B. and Daley, G.Q. (2013) Reprogrammed cells for disease modeling and regenerative medicine. *Annu Rev Med*, **64**, 277-290.
3. Check, E. (2002) A tragic setback. *Nature*, **420**, 116-118.
4. Marshall, E. (1999) Gene therapy death prompts review of adenovirus vector. *Science*, **286**, 2244-2245.
5. Mussolino, C. and Cathomen, T. (2012) TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol*, **23**, 644-650.
6. Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, **14**, 49-55.
7. Sun, N. and Zhao, H. (2013) Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol Bioeng*, doi: 10.1002/bit.24890.
8. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.
9. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
10. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. and Zhang, F. (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun*, **3**, 968.

11. Streubel, J., Blucher, C., Landgraf, A. and Boch, J. (2012) TAL effector RVD specificities and efficiencies. *Nat Biotechnol*, **30**, 593-595.

12. Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T. *et al.* (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell*, **12**, 238-251.

13. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, **39**, 9283-9293.

14. Sun, N., Liang, J., Abil, Z. and Zhao, H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst*, **8**, 1255-1263.

15. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, **29**, 143-148.

16. Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E. *et al.* (2013) A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol*, **31**, 251-258.

17. Choi, S.M., Kim, Y., Shim, J.S., Park, J.T., Wang, R.H., Leach, S.D., Liu, J.O., Deng, C.X., Ye, Z. and Jang, Y.Y. (2013) Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. *Hepatology*, doi: 10.1002/hep.26237.

18. Osborn, M.J., Starker, C.G., McElroy, A.N., Webber, B.R., Riddle, M.J., Xia, L., Defeo, A.P., Gabriel, R., Schmidt, M., Von Kalle, C. *et al.* (2013) TALEN-based gene correction for epidermolysis bullosa. *Mol Ther*, doi: 10.1038/mt.2013.1056.

19. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, **30**, 460-465.

20. Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y. *et al.* (2012) *In vivo* genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114-118.

21. Carlson, D.F., Tan, W., Lillico, S.G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D.F., Long, C.R., Whitelaw, C.B. and Fahrenkrug, S.C. (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A*, **109**, 17382-17387.

22. Guo, J., Gaj, T. and Barbas, C.F., 3rd. (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol*, **400**, 96-107.

23. Kim, H., Um, E., Cho, S.R., Jung, C., Kim, H. and Kim, J.S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods*, **8**, 941-943.

24. Kim, Y., Kweon, J. and Kim, J.S. (2013) TALENs and ZFNs are associated with different mutation signatures. *Nat Methods*, **10**, 185.

25. Cade, L., Reyon, D., Hwang, W.Y., Tsai, S.Q., Patel, S., Khayter, C., Joung, J.K., Sander, J.D., Peterson, R.T. and Yeh, J.R. (2012) Highly efficient generation of

heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res*, **40**, 8001-8010.

26. Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S. and Zhang, B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol*, **29**, 699-700.

27. Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D. and Holmes, M.C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods*, **8**, 74-79.

28. Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P. and Paques, F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther*, **11**, 11-27.

29. Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773-782.

30. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

31. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.

32. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.

33. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R. and Joung, J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*, **31**, 227-229.

34. Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, **31**, 233-239.

35. Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol*, **25**, 786-793.

36. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection. *Nat Methods*, **8**, 765-770.

37. Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*, **25**, 778-785.

38. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, **2**, 31-34.

39. Andersson, S., Davis, D.L., Dahlback, H., Jornvall, H. and Russell, D.W. (1989) Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J Biol Chem*, **264**, 8222-8229.

40. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and

assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res*, **39**, e82.

41. Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C. and Rebar, E.J. (2010) A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol*, **649**, 247-256.

42. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.

43. Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y. and Yan, N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720-723.

44. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.

.

CHAPTER 5. DEVELOPMENT OF A SINGLE-CHAIN TALEN ARCHITECTURE

5.1. Introduction

Transcription-activator like effector (TALE) nucleases (TALENs) have been widely used for genome editing or engineering in various organisms during the last three years (as reviewed in (1,2)). The long DNA binding sequences of TALE DNA binding domains enable TALENs to target the custom-selected loci precisely with minimal off-target effects or cyto-tyxociteis. The simple one-repeat-one-nucleotide DNA recognition code of TALE DNA binding domains provides modular DNA reading, which enables researchers to target essentially any genomic region.

The standard TALEN architecture is composed of TALE DNA binding domain, truncated N- and C-terminal extension and the non-specific DNA cleavage domain from a Type IIS restriction enzyme FokI (**Figure 5.1A**). Because the dimerization of the FokI cleavage domain is necessary for DNA cleavage activity (3), standard TALEN architecture requires two TALENs to bind to two DNA recognition regions flanking an unspecific central spacer. The optimal spacer length is dependent on the TALE scaffold construction (4-6). When the TALEN technology was initially developed, three limitations have been noticed to prevent its broader applications: a) the requirement of a 5'-thymidine preceding the TALE binding sites can limit the targeting space; b) The highly repetitive nature of the TALE DNA binding domain makes it difficult to synthesize TALEN coding genes using traditional cloning strategy; c) The bulky size of

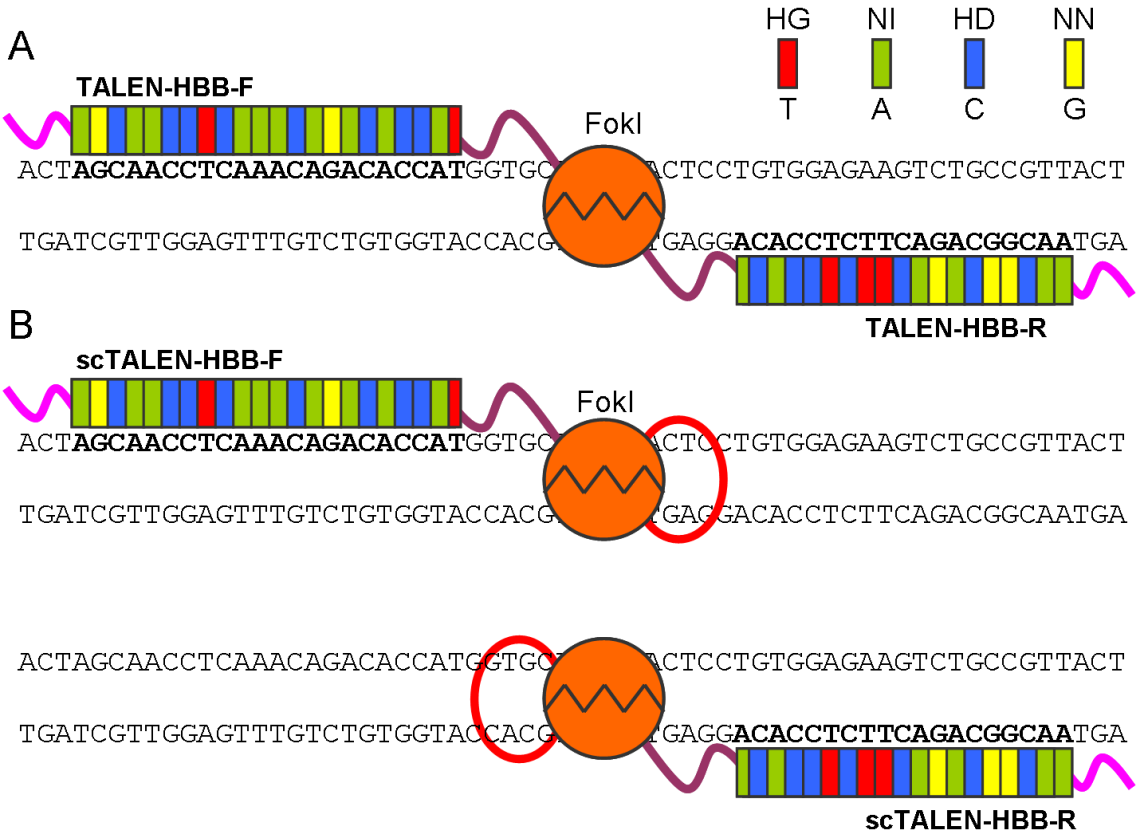

Figure 5.1. Schematic of standard and single-chain TALEN architectures. (**A**) Schematic of standard TALEN architecture. NTSs are represented by pink lines. CTSs are represented by magenta lines. TALEN binding sites are bold. (**B**) Schematic of scTALEN architecture. NTSs are represented by pink lines. CTSs are represented by magenta lines. TALEN binding sites are bold. A polypeptide linker that connects two FokI domains is represented by a red line.

TALENs possess the challenge of efficiently delivering corresponding DNA, RNA or proteins into certain cell types. In **Chapter 2**, we demonstrated that the requirement of the 5'-T can be mitigated by using a specific TALEN architecture with 63 amino acid C-terminal extension (6). Other studies also provided evidence that a thymine at position 0 is not strictly required for TALEN activity (4,7,8), indicating that the first limitation has been successfully addressed. The problem of synthesizing highly repetitive TALE genes

has also been solved by multiple groups by the introduction of Golden Gate cloning, solid-phase ligation and ligation-independent cloning techniques (7,9-14).

In order to address the third limitation, i.e. reduce the TALEN size, Beurdeley and coworkers have developed a compact TALEN (cTALEN) architecture by replacing the requisite dimeric FokI cleavage domain with the DNA cleavage domain from the I-TevI homing endonuclease (15). In this design, cTALENs can cleave the targeted DNA sequence as a monomer and thus reduce the TALEN payload to half as compared with the conventional TALEN architecture. However, unlike the non-specific FokI DNA cleavage domain, the I-TevI DNA cleavage domain has its own DNA recognition sequence. They discovered that the efficient DNA cleavage of cTALEN is dependent on the consensus sequence CDDHGS (D= A, G or T; H= A, C or T; S= C or G), which is outside the TALE binding site. Therefore, the native DNA recognition specificity of I-TevI DNA cleavage domain severely decreases the targeting space of cTALENs. In this chapter, I describe the development of a single-chain TALEN (scTALEN) architecture, in which two FokI DNA cleavage domains are linked by a polypeptide linker (**Figure 5.1B**). The appropriate linker is isolated from the high-throughput screening of a polypeptide linker library. In this design, the two FokI domains on the same polypeptide can form a catalytic active self-dimer and thus make scTALENs to function as monomers. The scTALEN architecture I have developed reduced the TALEN payload to half as compared to standard TALEN architecture. Moreover, because the FokI DNA cleavage domains have no DNA recognition specificity, the DNA target sequence is solely determined by the TALE DNA binding sites. Therefore, scTALENs can be used to target essentially any DNA sequence.

5.2. Results

5.2.1. Construction of a scTALEN library

Since FokI dimerization is required for DNA cleavage activity (3), we set out to construct a scTALEN by linking two FokI DNA cleavage domains with a polypeptide linker. Presumably the length and amino acid composition is critical to the formation of the intra-molecular FokI dimer. Because our knowledge is still limited for designing a functional linker *de novo*, we decided to apply the principle of directed evolution and screen one functional linker from a polypeptide linker library. The linker library is composed of various polypeptide linkers ranging from 30 to 120 amino acids with different sequences as previously described (16). At the beginning we generated a linker recipient expression vector with two FokI coding sequences separated by *BamH*I and *Xho*I restriction sites (**Figure 5.2A**). The library of polypeptide linkers were then inserted into the linker recipient expression vector to generate the library of scTALENs, in which the two FokI cleavage domains were joined by polypeptide linkers of varying length and amino acid composition. The complexity of the corresponding library is approximately 2×10^5. Each linker shared two common flanking polypeptides (GSGSGS) at both N- and C- termini. Theses Gly-Ser repeats were incorporated during construction of the linker library to provide flexible regions flanking linkers in order to minimize disruption of well-ordered domains or folds. The N-terminal Gly-Ser (*Bam*HI site) and C-terminal Gly-Ser-Ser (including *Xho*I site) provided additional flexibility for further manipulation or linker transfer (**Figure 5.2A & C**).

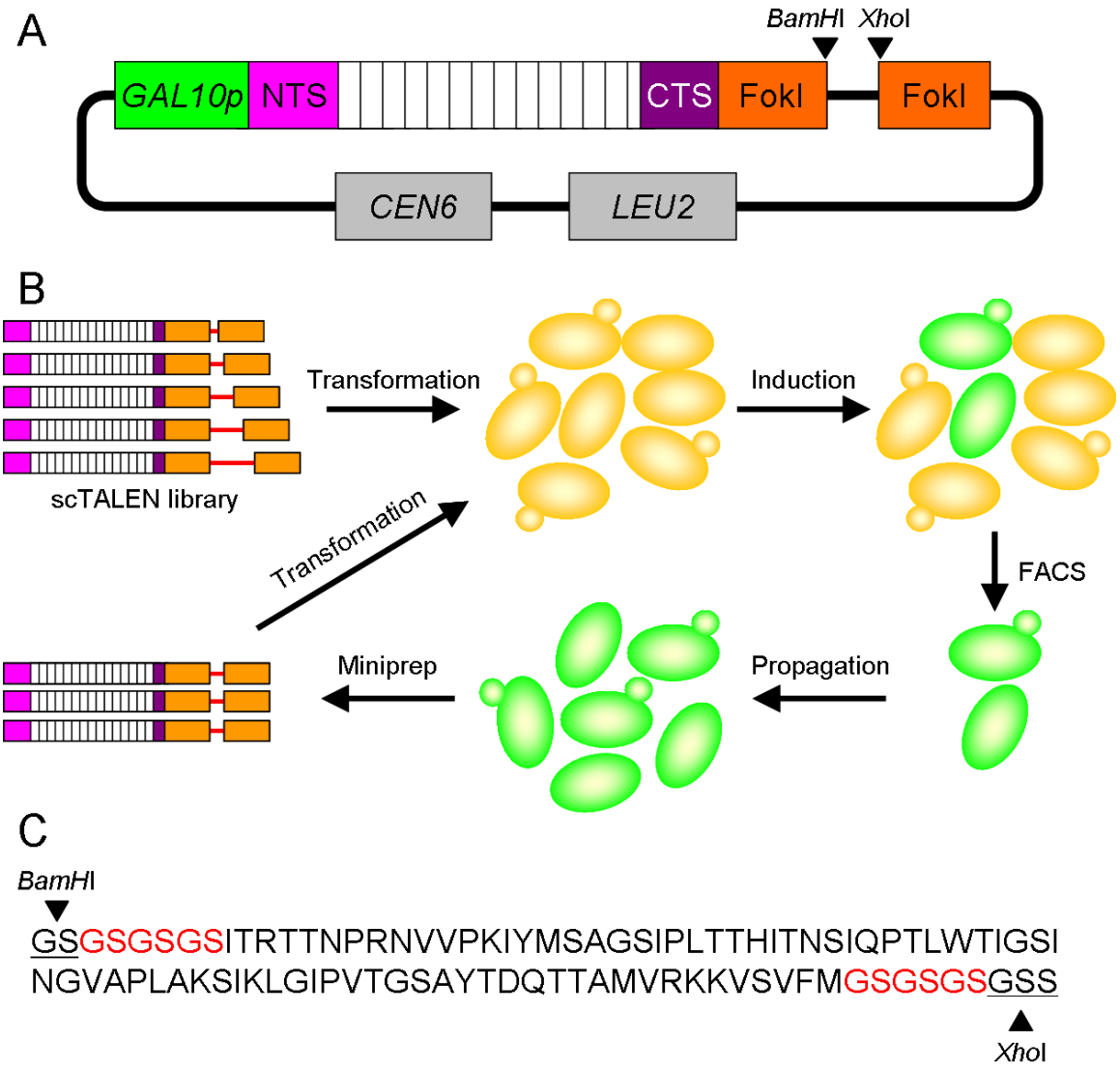

Figure 5.2. High-throughput screening of active scTALENs. (**A**) Schematic of a recipient expression vector. The peptide linkers with various lengths and sequences were cloned using the *BamH*I and *Xho*I sites. *GAL10p*, promoter of *GAL10* gene; *CEN6*, a centromere of yeast chromosome. *LEU2* auxotrophic marker is used for selection of transformants in yeast. (**B**) Schematic representation of the high-throughput screening in yeast for isolating active scTALENs. (**C**) Sequence information of the peptide linker from the isolated scTALEN. Common flanking polypeptides are shown in red. The N-terminal Gly-Ser (*BamH*I site) and C-terminal Gly-Ser-Ser (including *Xho*I site) are underlined.

5.2.2. High-throughput screening of active scTALENs

To identify catalytically active scTALENs in the library, we applied a high-throughput screening system in yeast as described in **Chapter 4**. This system couples the catalytic activity of a scTALEN with the enhanced green fluorescence protein (eGFP) signal (**Figure 4.1**). The expression of scTALEN was induced by galactose and the GFP-positive cells were collected by fluorescence-activated cell sorting (FACS). The sorted cells were pooled together and plasmids from the pool were isolated. The plasmids obtained from yeast cell miniprep were retransformed to *E. coli* and extracted again to increase plasmid quantity and quality. This process was continued in an iterative fashion until the scTALENs with catalytic activity were enriched and identified (**Figure 5.2B**). After three rounds of enrichments, we were able to identify one scTALEN architecture with two FokI domains linked by a polypeptide with 95 residues (**Figure 5.2C**). To characterize the isolated scTALEN architecture, we compared its activity with the standard TALEN architecture and a homing endonuclease I-CreI in the LacZ-based yeast reporter system as described in **Chapter 2**. The cleavage activities of these DNA endonucleases were indicated by the corresponding β-galactosidase activities (**Figure 2.2**). Whereas the conventional TALEN has no activity as a monomer, the scTALEN we isolated exhibited 13% activity as compared with I-CreI (**Figure 5.3**). To rule out the possibility that the obtained activity is from the evolution of the vector during high-throughput screening, we sub-cloned the scTALEN gene into a new expression vector. The recloned scTALEN expression plasmid showed similar activity as compared with the original scTALEN expression plasmid, confirming that the DNA cleavage activity is from the scTALEN open reading frame (**Figure 5.3**).

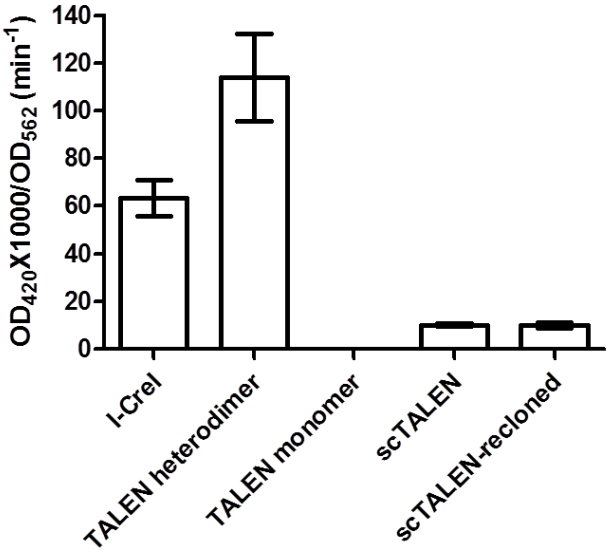

Figure 5.3. Activity of scTALEN in yeast cells as measured by the β-galactosidase assay (schematic of the assay is shown in **Figure 2.2**). Error bars indicate standard deviation of at least three independent experiments.

5.2.3. *In vivo* activity in human cells

To test the *in vivo* activity of the isolated scTALEN in human cells, we measured the genome editing efficiency in the modified surrogate reporter system as described in **Chapter 4**. The reporter construct encoding a monomeric red fluorescent protein (mRFP)-eGFP fusion protein was stably integrated into the genome of Hela cells. A TALEN target site within sickle human β-globin (HBB^S) gene locus (**Table 4.1** and **Figure 5.1**) (6) was inserted between the *mRFP* and e*GFP* genes, making the *eGFP* gene out of frame. In response to the TALEN-induced DNA cleavage, certain insertions and deletions caused by non-homologous end joining will make the *eGFP* gene in frame and restore eGFP function. Pairwise comparison was performed between conventional TALEN monomers and scTALEN architectures bearing the HBB-F and HBB-R DNA

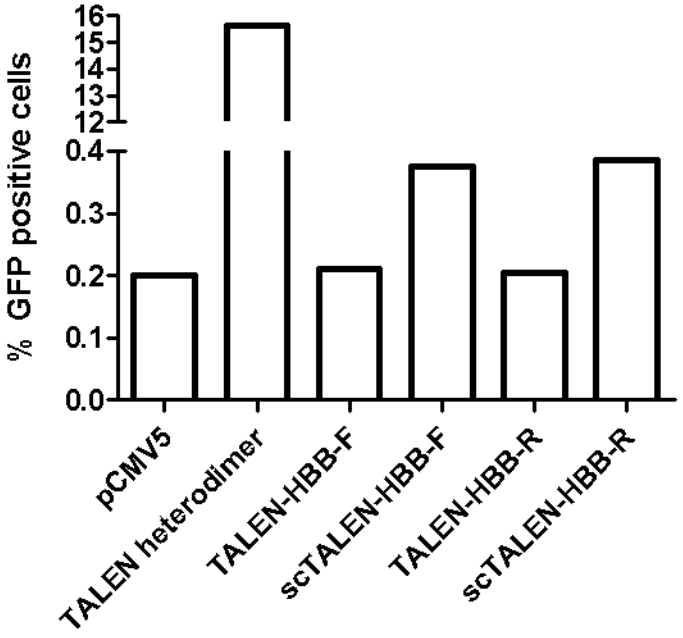

Figure 5.4. Activity of scTALENs in human cells as tested in the modified surrogate reporter system (schematic of the assay is shown in **Figure. 4.6**).

binding domains (**Figure 5.1**). The cells transfected with conventional TALEN monomers exhibited minimal GFP signals which were almost identical to background noise (**Figure 5.4**). However, the scTALEN monomers increased the GFP signal by over twofold. Although the scTALENs showed only 1-2% activities as compared to standard TALEN heterodimers, the scTALEN architecture we identified demonstrates an important proof-of-concept that FokI-based single-chain TALEN can be active in human cells. The scTALEN were then applied to target three endogenous sites of human genome with modest but significant activities (**Figure 5.5**). Western blot analysis indicates that the polypeptide linker and another FokI DNA cleavage domain in the scTALEN architecture do not change protein expression levels in human cells (**Figure 5.6**).

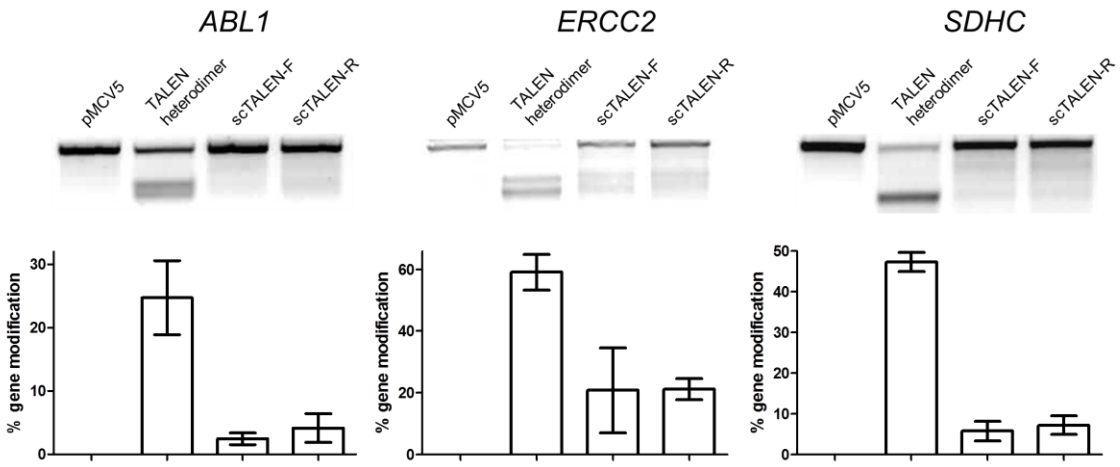

Figure 5.5. Application of the scTALEN architecture for human genome editing. SURVEYOR nuclease assay (procedure is described in **4.4.8**) was performed to determine the frequencies of scTALEN-induced gene modifications.

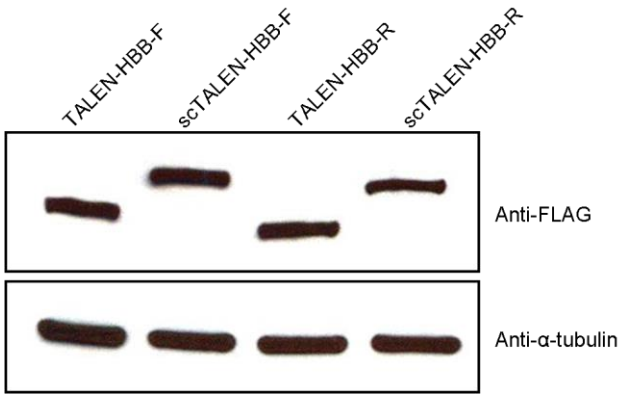

Figure 5.6. Western blot analysis for the expression of TALEN variants. All the TALEN variants contained a FLAG tag at the N-terminus. α-tubulin was used as an internal control.

5.3. Discussion

In this study, I developed a scTALEN architecture by fusing two FokI DNA cleavage domains at the C-terminus of the TALE DNA binding domain. A similar strategy has been applied for the construction of single-chain zinc finger nucleases (17,18). However,

the flexible $(GGGGS)_n$ linkers used in the two studies were not functional for the construction of scTALEN architecture (data not shown). To facilitate the dimerization of the two FokI domains on the same polypeptide chain, an appropriate linker with certain length and secondary structure is critical. Because the tools are still lacking for the design of a functional linker *de novo*, a directed evolution strategy was harnessed to isolate a functional peptide linker from a library of linkers with various lengths and sequences.

The scTALEN architecture I have developed showed only moderate *in vivo* activity in both yeast and human cells. Therefore, further studies are needed to improve its genome editing efficacy. Using this scTALEN as a template, I created another scTALEN library by the introduction of random mutations on the FokI-linker-FokI region. Three rounds of directed evolution were carried out using the same high-throughput screening system. However, no mutant with increased activity can be isolated, indicating the limitation of the screening system. A detailed structural analysis of the 95 amino acid polypeptide linker can provide a detailed mechanism of the formation of an intra-molecular FokI dimer. This structural information can also be helpful in the design of more active linkers using rational protein design. Despite the low *in vivo* activity, the scTALEN architecture I have developed represents the first example of single-chain FokI-based TALEN, which represents a very important proof-of-concept that a FokI dimer can be formed in a single polypeptide chain and paves the road for future protein engineering endeavors.

5.4. Materials and Methods

5.4.1. Materials

Q5 DNA polymerase, T4 DNA ligase, Antarctic phosphatase, and restriction

endonucleases were purchased from New England Biolabs (Beverly, MA). QIAprep Spin Plasmid Miniprep Kit, QIAquick Gel Extraction Kit, and QIAquick PCR Purification Kit were obtained from Qiagen (Valencia, CA). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). All the other reagents unless specified were obtained from Sigma-Aldrich (St. Louis, MO).

5.4.2. Yeast *eGFP* reporter strain

The single-strand annealing eGFP reporter on pRS414 vector with two separated *eGFP* fragments sharing a 100 bp homologous region was described in **Chapter 4**. The centromere sequence of the plasmid was then removed by the *Pml*I digestion and blunt ligation. This plasmid was then digested with *Hind*III within the *TRP1* gene. The linearized plasmid was transformed into *S. cerevisiae* HZ848 (*MATα*, *ade*2-1, *ade*3Δ22, *Δ*ura3, *his*3-11,15, *trp*1-1, *leu*2-3,112 and *can*1-100) using the LiAc/SS carrier DNA/PEG method (19) and stably integrated into the chromosome at the *TRP1* site. The yeast reporter strain was obtained by selection of the transformed yeast cells on plates containing synthetic complete medium lacking tryptophan with 2% glucose at 30 °C. The integration was confirmed by PCR amplification and DNA sequencing of the targeted genomic region.

5.4.3. Creation of a scTALEN library

The yeast TALEN expression vector with pRS415 backbone was described in **Chapter 2**. This vector was digested with *Hind*III and *Sal*I to remove the latter part of the FokI coding sequence. Ligation of this digested vector and a PCR product amplified by the

primers FokI-middle(HindIII)-for 5'-gga tac t*aa agc tt*a tag cgg agg-3' (*Hind*III site shown in italic) and SalI-XhoI-BamHI-FokI-rev 5'-atc cgc gtc gac act gac tga ctg act cga gcc tac tga ctg tta *gga tcc* aaa gtt tat ctc gcc gtt'3' (*Xho*I site underlined and *Bam*HI site shown in italic) digested with *Hind*III and *Sal*I makes plasmid pRS415-GAL10P-N3-C1-Xa10-BamHI-XhoI. An additional copy of FokI coding region was inserted into this vector between the *Xho*I and *Sal*I sites to generate the recipient expression vector (**Figure 5.2A**).

The polypeptide linker library was kindly provided by Dr. David Baker from the University of Washington (Seattle, WA). The sequences encoding the linker library was PCR amplified by primers: BamHI-linkerlib-forN 5'-act gac *gga tcc* ggt agc ggc tca gga-3' (*Bam*HI site shown in italics) and XhoI-linkerlib-revN 5'-act gac *ctc gag* cc g ctt ccc gac cca ga-3' (*Xho*I site shown in italics). The PCR product has various lengths and showed a smeared wide band on 2% agarose gel. All of the PCR products were gel-extracted, digested with *Xho*I and *BamH*I, purified, and inserted into the recipient expression vector that was digested with the same restriction enzymes. The ligation products were precipitated by 1-butanol and electroporated into DH5α competent cells to create the library of scTALENs. The library contains approximately 2×10^5 transformants (data not shown).

5.4.4. High-throughput screening

The plasmids were minipreped from the scTALEN library and transformed into the reporter yeast strain as described above using the LiAc/SS carrier DNA/PEG method (19). After transformation, the yeast cells were recovered at 30 °C with shaking for 1 h in YPA

medium (1% yeast extract, 2% peptone and 0.01% adenine hemisulphate) with 2% glucose. The cells were then centrifuged and resuspended in synthetic complete medium lacking leucine and tryptophan (SC-Leu-Trp) with 2% raffinose for growing at 30 ℃ with shaking overnight. TALEN expression was induced by culturing cells in YPA medium with 2% galactose at 30 °C for 5 h. The cells were then centrifuged and resuspended in SC-Leu-Trp liquid medium with 2% glucose at 30 ℃ with shaking overnight. On the next day, cells were analyzed on a BD FACS Aria III cell sorting system (BD Biosciences, San Jose, CA) and 20,000-100,000 eGFP-positive cells were collected. After 2 days' growth at 30 °C in SC-Leu-Trp liquid medium with 2% glucose, plasmids from the sorted cells were extracted using Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research, Orange, CA). The plasmids were then electroporated into DH5α competent cells for amplification. The minipreped plasmids were subjected to the next round of screening.

5.4.5. *LacZ* reporter assay in yeast

The yeast reporter system based on *lacZ* gene was described in **Chapter 2**. To test the *in vivo* cleavage activity of scTALENs in the yeast reporter system, both the pRS414 reporter plasmid and pRS415 expression plasmid were transformed into *Saccharomyces cerevisiae* HZ848 (*MATα*, *ade*2-1, *ade*3Δ22, *Δ*ura3, *his*3-11,15, *trp*1-1, *leu*2-3,112 and *can*1-100). After 3-4 days growth at 30 °C on SC-Leu-Trp plates with 2% glucose, three transformants were picked to grow in SC-Leu-Trp liquid medium with 2% galactose. After cultivation for 2 days at 30 °C, yeast cells were harvested by centrifugation and lysed by Y-PER yeast protein extraction reagent (Pierce, Rockford, IL). The β-

galactosidase (β-GAL) enzyme activity was determined by the β-galactosidase enzyme system from Promega (Madison, WI) according to the manufacturer's instruction. For each sample, the β-GAL activity was measured as a function of time and normalized by total protein concentration determined using the BCA protein assay (Pierce, Rockford, IL). Homing endonuclease I-CreI (20) and conventional dimeric TALENs served as positive controls.

5.4.6. *In vivo* activity in human cells

The pCMV5 TALEN expression plasmids were reported in **Chapter 2**. To construct scTALEN expression vectors for human cells, the FokI-linker-FokI coding sequence of scTALEN was PCR amplified from the pRS415 plasmids. The PCR products were then digested with *Afl*II and *Sal*I and inserted into the *Afl*II/*Sal*I digested pCMV5 TALEN expression plasmids to make pCMV5 scTALEN expression plasmids by replacing the FokI genes with the FokI-linker-FokI coding sequences.

A modified surrogate reporter system in human cells was described in **Chapter 4**. Reporter Hela cells were routinely maintained in the modified Eagle's medium (MEM) supplemented with 10% Fetal Bovine Serum (FBS; Hyclone, Logan, UT). Cells were seeded in 24-well plates at a density of 5×10^4 per well. After 24 h, reporter cells were transfected with 1 μg of pCMV5 scTALEN expression plasmids using FuGene HD transfection reagent (Promega, Madison, WI) under conditions specified by the manufacturer. Cells were trypsinized from their culturing plates 48 h after transfection and resuspended in 300 μL phosphate buffered saline (PBS) for flow cytometry analysis. 20 000 cells were analyzed by a BD LSRII flow cytometer (BD Biosciences, San Jose,

CA) to quantify the eGFP-positive cells.

5.4.7. Western blot analysis

Hela cells in 12-well plates were harvested 24 h after transfection with TALEN expression plasmids. Cells were collected by centrifugation, washed with PBS and resuspended in 40 μL whole cell lysis buffer (1 M Tris-HCl, pH 6.8, 20% sodium dodecyl sulfate, and 0.1 M dithiothreitol). Proteins were resolved by 4-20% Mini-Protean TGX Precast Gel (Bio-Rad Laboratories, Hercules, CA), transferred onto a nitrocellulose membrane, blocked for 1 h with Tris-buffered saline/0.05% Tween 20 containing 5% nonfat milk, followed by incubation with anti-FLAG tag (1: 500) and anti-α-tubulin (1: 15 000) antibodies at 4 °C overnight. After incubation with anti-mouse horseradish peroxidase-conjugated secondary antibody (1: 15 000; GeneScript, Piscataway, NJ) for 1 h, bands were visualized using SuperSignal West Pico Chemiluminescent Substrate (Pierce, Rockford, IL).

5.5. References

1. Sun, N. and Zhao, H. (2013) Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol Bioeng*, doi: 10.1002/bit.24890.
2. Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, **14**, 49-55.

3. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10570-10575.

4. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, **29**, 143-148.

5. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, **39**, 9283-9293.

6. Sun, N., Liang, J., Abil, Z. and Zhao, H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst*, **8**, 1255-1263.

7. Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P. and Church, G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res*, **40**, e117.

8. Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., Feng, J., Verdier, V. and Szurek, B. (2011) Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae pv. oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Mol Plant Microbe Interact*, **24**, 1102-1113.

9. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res*, **39**, e82.

10. Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T. *et al.* (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell*, **12**, 238-251.

11. Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E. *et al.* (2013) A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol*, **31**, 251-258.

12. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, **30**, 460-465.

13. Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol*, **31**, 76-81.

14. Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, **6**, e19722.

15. Beurdeley, M., Bietz, F., Li, J., Thomas, S., Stoddard, T., Juillerat, A., Zhang, F., Voytas, D.F., Duchateau, P. and Silva, G.H. (2013) Compact designer TALENs for efficient genome engineering. *Nat Commun*, **4**, 1762.

16. Li, H., Pellenz, S., Ulge, U., Stoddard, B.L. and Monnat, R.J., Jr. (2009) Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res*, **37**, 1650-1662.

17. Minczuk, M., Papworth, M.A., Miller, J.C., Murphy, M.P. and Klug, A. (2008) Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. *Nucleic Acids Res,* **36**, 3926-3938.

18. Mino, T., Aoyama, Y. and Sera, T. (2009) Efficient double-stranded DNA cleavage by artificial zinc-finger nucleases composed of one zinc-finger protein and a single-chain FokI dimer. *J Biotechnol*, **140**, 156-161.

19. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, **2**, 31-34.

20. Chevalier, B.S., Monnat, R.J., Jr. and Stoddard, B.L. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol*, **8**, 312-316.